INVESTIGATIONS ON THE PROPERTIES AND ESTIMATION OF EARTH RESPONSE OPERATORS FROM EM SOUNDING DATA

by

Andreas Tzanis

Thesis presented for the degree of
Doctor of Philosophy of the
University of Edinburgh
1987

DECLARATION

I hereby declare that the work presented in this thesis is my own unless otherwise stated in the text, and that the thesis has been composed by myself.

Andreas Tzanis

Da steh' ich nun, ich armer Tor
und bin so Klug, als wie zuvor....

Acknowledgements

The compilation of this thesis would never have been possible, without the untiring encouragement and support, material and moral, that I received from my parents. They have always been there, when they were needed, with patience, understanding and care. I do not know whether my gratitude will ever be enough to match their efforts.

This thesis was supervised by two persons. I have been very fortunate to work with Dr D. Beamish, in his EM induction group in the British Geological Survey. He has always been willing to discuss and analyse with me the problems encountered throughout my investigations, always with an open mind, and, a lot of patience towards a demanding person, as I know myself to be; he is hereby recommended to future MT student generations, as the supervisor and friend to look for. I was also fortunate to be a student of Dr. V.R.S. Hutton, of the University of Edinburgh; I received continuous encouragement and assistance from a supervisor who is always eager to deal with the problems of her students.

I am indebted to the people of the Geomagnetism Research Group of the British Geological Survey for their hospitality; I would like to thank them, each one and all together for providing a warm and friendly place to work in, for four uninterrupted years. I only regret that John Riddick and I did not settle the hot question of whom certain marbles belong to, as yet; I can only hope that this will be done at a later time.

I acknowledge several discussions with Dr B. Hobbs, that proved very useful in tackling some problems related to my investigations. I would also like to thank Dr R. Hipkin, for his assistance in matters related or unrelated to this thesis. Last, but not least, I would like to thank Prof. and Mrs K. Creer for their friendliness, when most needed, and throughout my stay in Edinburgh.

ABSTRACT

This thesis comprises 6 separate studies, of problems associated with the Magnetotelluric and Audio-magnetotelluric crustal sounding method, arranged in the form of four chapters and one appendix. Following an introduction to the thesis, the first chapter is concerned with the problem of recovering the geoelectric structural directions and dimensionality, from the observed magnetotelluric Earth response tensors (operators), using a group theoretical approach. The Earth response ope- operators are described with a SU(2) symmetry, which permits their analysis in three spatial dimensions, and provides an efficient and precise parameter estimation procedure. An application of the theory to actual magnetotelluric data is presented and discussed. The second chapter comprises two parts. The first part is a discussion of the magnetotelluric response operator theory, from a time domain perspective. Some relationships derived from the time domain considerations are compared with existing (frequency domain) theory, and the implications of the comparisons are pointed out and discussed. The second part, comprises a spectral study of magnetotelluric field (data) and noise interactions, using the poweful high resolution Maximum Entropy (MAXENT) spectral analysis method, with emphasis on cultural (anthropogenic) noise interference. The results are compared to conventional (FFT based) estimation procedure. The comparisons allow some useful observations to be made, about distortions that are undetectable by conventional estimation methods; these are pointed out and discussed in detail. A method to smooth distorted high resolution response functions is also presented. Chapter 3 is concerned with the problem of eliminating the harmonic ('50Hz') part of the of the anthropogenic noise; a number of potentially useful methods are presented and discussed. Chapter four is concerned with the problem of eliminating the random part (transient stray currents) of anthropogenic noise. A spectral procedure, based on MAXENT, that involves the analysis of short data series is discussed. Two time domain data screening

methods are also introduced, and examples of their performance are given. Considerable attention is given to robust statistical estimation techniques, in both the time and frequency domains, and an easy-to-implement robust statistical approach is described and discussed. Appendix I contains published investigations on the time-local behaviour of an important class of source fields, the Schumann resonances of the Earth-ionosphere cavity. Both their spectral and polarization characteristics are investigated, together with their usefulness in Audiomagnetotelluric crustal sounding.

CONTENTS

CHAPTER 3
ON THE PROBLEM OF ELIMINATING THE HARMONIC (50Hz) INTREFERENCE FROM THE POWER DISTRIBUTION GRID

CHAPTER 4
ON THE PROBLEM OF DATA QUALITY EVALUATION, SELECTION AND PROCESSING

INTRODUCTION

Αει ο θεος ο μεγας γεωμετρει...

The Magnetotelluric (MT) method has come a long way since the time of its introduction by Cagniard (1953). Some research workers would even say that it has reached maturity, an opinion however that is not shared by this author. Several, quite basic problems are far from having found a solution yet; these include some quite fundamental concepts. The author's favourite example is the one-dimensional inverse problem, which many authors (in particular authors of inversion schemes), will consider 'largely solved', but for which a basic theoretical consensus and understanding has not been reached as yet (for example, see the review of Parker, 1983). In the meantime, new one dimensional inversion procedures continue to appear, while credible two- and three-dimensional inversion schemes do not exist as yet. In addition, the MT method is somewhat 'handicapped', because it attempts to establish the geoelectric structure from measurements of the horizontal Electromagnetic (EM) field components only. It is virtually impossible to measure the vertical electric field component, which introduces severe limitations in the case of EM induction over three-dimensional geoelectric structures; this is just another problem we all find it very convenient to forget... The question of reliable data acquisition, especially in the case of MT field contamination by alien EM fields (noise), natural or anthoropogenic, poses a problem that we all are aware of. This is but a short list of the problems that are still with us and require our attention.

The present work will attempt to look into some of these problems, and

hopes to contribute to their better understanding. The focus of the investigations is a number of topics regarding the theory and estimation of MT response functions that include:

(a) The basics of the MT problem, from a time domain point of view.

(b) The problems of recovering the geoelectric structural directions and dimensionality.

(c) The problems of understanding the behaviour of EM induction data recorded in intense EM noise domains, and their recovery, with particular application to the sub-audio and audio frequency range (AMT).

(d) Understanding the behaviour of an important class of source fields, i.e. the sub-audio and audio-frequency Schumann resonances of the Earth- ionosphere cavity.

The topics treated in this work are diverse, and as a consequence, diverse analysis techniques and methodologies, borrowed from different fields of science and statistics have been implemented. A detailed description of all these methods would, be at least difficult, if not unwarranted. Thus, only essential information is provided; for more details, the reader is kindly requested to refer to the bibliography list.

This volume comprises four chapters and one appendix. Each of these units contains (roughly) investigations on one of the topics above, and, as a consequence they comprise independent studies within the framework of MT. Chapter 1 concentrates on problem (b). The most important parameters we seek to obtain with the MT sounding method are, the magnitude of energy dissipation experienced by the source field, i.e. the admittance/impedance (response) functions of the Earth, and, the spatial directions of the geoelectric structure producing them. These parameters are interrelated and contained in the observation; the

response functions extremize in association with the geoelectric structural directions. Moreover, they are associated with, and allow inferences about, the dimensionality of the geoelectric structure. The conventional method used to unveil the geoelectric structure is described in Part 1 of this chapter, and comprises a transformation (rotation) of the experimental frame of reference, so as to coincide with the structural directions. This method however suffers from severe limitations that render it more problematic than is usually appreciated. The first to demonstrate the problems was Eggers (1982), who, in addition, suggested a more reliable method of MT parameter estimation. His work appears to have initiated a period of intense research on the subject, and several publications appeared, that point out shortcomings in Eggers' method and suggest alternative procedures.

Part 2 of this chapter attempts to provide a comprehensive solution to the problem, through an analysis of the three-dimensional (3D) MT response functions with the tensors of the three-dimensional pure rotation group SU(2), of 2x2 unitary matrices. A description of the properties of the rotation group is provided. The application of SU(2) is shown to amount to a decomposition of the observable 2x2 MT tensors into their spatial components (geoelectric structural directions), and magnitude components (extremal values equivalent to eigenvalues), associated with energy dissipation (i.e. with the conductivity/resistivity of the Earth) along the structural directions. The analysis shows that in the case of a 3D geoelectric stucture, a 3D description of the MT tensors is possible; the structural directions may lie at arbitrary positions in the 3-space, and are not confined to the horizontal plane, as most analyses appear to suggest. This 3D information is recoverable, if we treat the observations with the

matrices of SU(2). It is also shown that the analysis readily reduces to the limiting (two-dimensional) case of geoelectric structure. The application of the method is compared to the results of other approaches, a review of which is provided together with the comparisons. It is shown that the method locates the extremal values of the MT tensors, within a few parts in 10^{-8}, a performance which is superior to that of other methods. The theoretical review suggests that some of the other methods are, in fact, partial cases of the proposed generalized analysis scheme.

Part 3 of the first chapter is an application of the method to actual MT data, from N. England and S. Scotland, kindly provided by Dr D. Beamish. The analysis shows that the proposed method is consistent with expectations for the limiting (one- and two-dimensional) cases. Moreover, it emphasizes the importance of correct coordinate frame transformations in the two- and three-dimensional cases, by pointing out that conventional practices may return parameters that display 'order of magnitude' differences with respect to the true extremal values. The analysis reveals several data features that are undetectable by conventional procedures. These include evidence of influences, on local MT soundings, by regional induction processes; the MT data properties thus revealed, appear to be consistent with the results of regional induction studies published by Banks et al (1983) and others.

Chapter 2 comprises two independent parts. The second part however, makes use of some theoretical concepts developed in the first. Thus, it was decided that they could be presented together. Part 1 is a discussion on the nature of the Earth response operators (functions) from a time domain perspective. The discussion begins with a

description of some fundamental properties of physical (realizable) time domain operators, and specializes in the one-dimensional case of geoelectric structure, comprising a stack of uniform conducting slabs. The response function is constructed in a manner reminiscent of the analogous functions in reflection seismology. This approach provides some insights into the properties of an Earth response operator, with some interesting implications, concerning the uncertainties involved in our observations, and, most importantly, the space of solutions that is fundamental to the inverse MT problem. With respect to the latter, the discussion appears to support Parker's (1980) conclusions. A theoretical analogy with the one dimensional reflection seismology problem is pointed out. Finally, a presentation of the analytical properties of the multidimensional extension of the time domain response function is provided.

Part 2 is a study of the interaction of MT fields and noise processes, with emphasis on the anthropogenic (cultural) interference. The study is conducted entirely in the frequency domain. The details of the data and noise interactions (signal to noise ratio), are studied by constructing high (frequency) resolution Earth response functions, using the Maximum Entropy (MAXENT) spectral analysis method. MAXENT is used in its multidimensional (multivariate) form, and a discussion on the implementation and rationale of the method is provided as an Appendix to the Chapter. The high resolution frequency domain response functions are compared with the conventional (discrete) response functions afforded by the Fast Fourier Transform (FFT) method. It is shown that for certain types of anthropogenic noise (unstable power distribution grid), harmonic and subharmonic generation may produce distortion over a considerable portion of the bandwidth, which, in extreme cases, may

extend over a few decades. It is also shown that in some cases, the stabilization (discretization by band-avaraging) procedure incorporated in the FFT approach, may lead to grossly erroneous response function estimates, albeit with coherent characteristics. A discussion on the generation and received characteristics of noise waveforms due to an unstable power distribution grid is also included.

In case one decides to use the high resolution spectral analysis procedures, described in Chapter 2, the final section provides a technique to smooth the instabilities that may be observed in the resulting high resolution Earth response function. This is based on a moving summation (convolution), in the frequency domain, of the observed response with an exponentially decreasing function of frequency, and appears to be consistent with the analytical properties of an MT operator. Although it must be viewed as an interim solution, the method appears to be quite effective.

Chapter 3 is dedicated to the problem of eliminating the harmonic (predictable) part of anthropogenic interference. Emphasis is again placed on noise waveforms emanating from an unstable power distribution grid. The performance of an established noise reduction method (Delay Line Filtering, Fischer 1982) is evaluated, and two more techniques are introduced and described. These methods are based on time series modeling and stochastic deconvolution. The idea behind these methods is to generate an optimal representation (model) of the noise process, and, subsequently, use it to remove the noise waveforms from the recorded data. The performance of these methods is demonstrated with a few examples.

The subject of Chapter 4 is the elimination of another form of

anthropogenic noise, i.e. the transient waveforms originating due to the switching of current loads, and are generally known as stray currents. These waveforms mainly affect the telluric data channels. The short duration and random occurence of these waveforms inspired an investigation into the time-local properties of noise and data interaction, with emphasis on the spectral characteristics of short data lengths. It is concluded that valid response functions can be generated by processing short data lengths; this would be advantageous in the case of contamination by stray currents.

The rest of Chapter 4 describes two novel methods for data screening (selection), with particular applicability in the presence of stray currents. It is shown that simple time domain functions can be constructed, that epitomize the relative data and noise content for any data window. These functions comprise:

(a) A multiple correlation coefficient between the magnetic and telluric channels, that may be used as an acceptance criterion, to supplement the conventional predicted coherence functions, and

(b) a scale factor that can serve as an indicator of the relative power content, and transfer of power between the magnetic and telluric channels. Data windows contaminated with transient waveforms display inconsistent power in the magnetic and telluric fields, and therefore erratic scale factors. It is shown that this information can be manipulated, so as to be useful in detecting and rejecting data realizations contaminated by stray currents. The manipulation and processing of the information contained in these scale factors is conducted with robust statistical techniques, and in particular, regression-M estimation (e.g. Huber, 1981). An assessment of various implementations of the regression-M estimation procedure is provided,

and their numerical performance is also evaluated. Subsequently, a robust statistical method is described, that appears to be suitable for the AMT data sets under consideration. An extension of these time domain robust regression-M procedures to frequency domain estimation exists, and is described. The effectiveness of the combination of joint short data length processing and robust statistical time domain screening, is also demonstrated. A discussion of the ideas presented in chapter 4 is also included.

The work contained in Appendix I was carried out for the purposes of the present thesis. It contains studies of the spectral and polarization properties of a very important class of (A)MT source fields, the Schumann resonances of the Earth-Ionosphere cavity, as well as an detailed evaluation of their usefulness in passive EM crustal sounding. However, the author was fortunate enough to see the work published prior to the completion of this volume. Thus, it was decided that the published work would be presented in its published form, partly to avoid duplication of existing material, and partly because the author believes that it is more presentable and readable in this way. Therefore, the reader is hereby requested to view the contents of Appendix I as an equivalent fifth chapter; it is not possible to label it as one, for purely technical reasons (thesis presentation regulations). The contents of the papers presented in Appendix I can be found in their associated abstracts, and will not be detailed here.

As stated above, this thesis comprises a study of the MT operators using several, and quite diverse techniques. There is, however, a common denominator to all the analysis methods implemented. They are, in one way or another, geometrical (or rather topological) analysis methods,

ranging from the tensor transformations of Chapter 1, to the robust statistical techniques of Chapter 4, which are but methods of descent in a topological continuum. Thus, I thought it appropriate to adopt as a motto, a 'mystical' mnemonic rule used by the late Pythagoreans; the Greek verse in page 1 says, roughly, "God almighty is forever a Geometrist". In my opinion, this verse can be viewed as the general idea behind the structure of our World - the Universe is Geometry, very much in accordance with the views of modern Physics.. If, however, you count the number of letters in each (Greek) word of the verse, you will end up with the first six digits of π...

CHAPTER 1

AN ANALYSIS OF THE MAGNETOTELLURIC OPERATORS IN THREE DIMENSIONS

PART 1:

THEORETICAL BACKGROUND

I. The one dimensional (1-D) Earth - passive induction in layered media

The goal of Magnetotelluric sounding is to derive the electrical conductivity structure of the Earth, from measurments of the Electromagnetic (EM) field components at its surface. The propagation and attenuation of EM waves in any conducting medium is described by Maxwell's equations

$$\text{curl}\mathbf{H} = (\sigma + i\omega\varepsilon)\mathbf{E} \qquad (1)$$

$$\text{curl}\mathbf{E} = -i\omega\mu_0\mathbf{H} \qquad (2)$$

$$\text{div}\mathbf{H} = 0$$

$$\text{div}\mathbf{E} = 0$$

where $\mathbf{E}$ and $\mathbf{H}$ represent the electric and magnetic field vectors respectively, σ is the conductivity of the medium, μ_0 is the magnetic permeability of free space, and, ε is the electric permittivity of the medium. By eliminating $\mathbf{H}$ or $\mathbf{E}$ from (1) and (2) one obtains the vector Helmholz equations

$$\text{curl.curl}\mathbf{E} + \mathbf{k}^2\mathbf{E} = -i\omega\mu_0\mathbf{J}_e \qquad (3a)$$

$$\text{curl.curl}\mathbf{H} + \mathbf{k}^2\mathbf{H} = \text{curl}\mathbf{J}_e \qquad (3b)$$

$$\mathbf{J}_e = i\omega\varepsilon\,\mathbf{E} \qquad (3c)$$

where $\mathbf{J}_e$ is the current density for the external source field (displacement current), and $\mathbf{k}^2 = i\omega\mu_0\sigma$ represents the wavenumber of the EM field. The solution of (3), depends on the assumptions made about the distribution of σ across the Earth, and the source fields considered. Cagniard's (1953) introduction of the MT problem is based

on the assumption of a horizontal plane-wave **H** field, propagating into the Earth, and inducing currents in a source-free electrically layered and laterally isotropic medium, i.e. one for which the conductivity σ varies only as a function of depth. Also, one finds that for the frequency range of interest, (f< 1000Hz), and, the rock conductivity ranges possible, the displacement currents J_e are negligible, because $\sigma >> \omega\varepsilon$ (e.g. Keller and Frischknecht, 1966). This is the quasi-static limit, where only conduction currents are of any importance. Given the above constraints, and assuming a Cartesian coordinate frame with x pointing North, y pointing East and z pointing into the Earth, all the partial derivatives of **E** and **H** with respect to x and y vanish, and (3) reduces to the homogeneous Helmholz equation

$$\begin{matrix} \partial^2\mathbf{E}/\partial z^2 \\ \partial^2\mathbf{H}/\partial z^2 \end{matrix} - k^2 \begin{bmatrix} \mathbf{E} \\ \mathbf{H} \end{bmatrix} = 0 \qquad (4)$$

where $\mathbf{E} = [E_x, E_y, 0]$ and $\mathbf{H} = [H_x, H_y, 0]$. Assuming an $\exp(-i\omega t)$ time dependence, the solutions of this type of equation have the form

$$\mathbf{F}(z) = (A \exp(kz) + B \exp(-kz)) \qquad (5)$$

where $\mathbf{F} = [\mathbf{E}, \mathbf{H}]$. The right part of the sum represents a vertically downgoing (diffusing), and the left part a vertically upgoing (scattered) wave. From the curl equations, the components of the two fields are found to be related by

$$\partial E_x/\partial z = i\omega\mu_0 H_y \qquad (6a)$$

$$\partial E_y/\partial z = -i\omega\mu_0 H_x \qquad (6b)$$

Combining (5) and (6), we find that the relationship between the EM field components can be writen as

$$\begin{bmatrix} E_x \\ E_y \end{bmatrix} = \frac{i\omega\mu_0 (A\exp(kz) + B\exp(-kz))}{k (A\exp(kz) - B\exp(-kz))} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} H_x \\ H_y \end{bmatrix} \qquad (7)$$

The ratio $Z = E_x/H_y = -E_y/H_x$ is the impedance of the medium and is a

scalar quantity that depends on the geoelectric properties of the Earth. In order to solve (4), in the case of a layered Earth with plane interfaces, one applies the usual boundary conditions

(a) The tangential component of **E** is continuous across the interface

(b) The tangential component of **H** is continuous across the interface

(c) The normal component of **H** is continuous across the interface.

These boundary conditions facilitate the elimination of the constants A,B and lead to the recursive formula (e.g. Keller and Frischknect, 1966)

$$Z(0) = -i\omega\mu_0/k \; [k_1h_1 + \text{arcoth}\{k_1/k_2 \coth(k_2h_2 + \text{arcoth}[k_2/k_3 \ldots$$
$$\ldots\text{arcoth}\{k_{n-2}/k_{n-1}\coth(k_{n-1}h_{n-1}+\text{arcoth}[k_{n-1}/k_n])\}\ldots])\}]$$

with k_j and h_j denoting the propagation constant (wavenumber) and thickness of the j'th layer respectively. Another characteristic of (7) is that the scalar product between the two fields, disappears, i.e. $\mathbf{E}.\mathbf{H} = 0$. This feature denotes the Transverse Electro-Magnetic (TEM) mode of wave propagation within the conducting layered Earth.

When the Earth medium becomes a uniform conducting half-space, the expression for the impedance reduces to $Z = E_x/H_y = i\omega\mu_0/k$, from which the resistivity of the half-space can be determined from the fields by

$$\rho_a = 1/\sigma = (\omega\mu_0)^{-1}|E_x|^2/|H_y|^2 = (\omega\mu_0)^{-1}|Z|^2$$

which, for the practical units of mV/Km, nT, and cyclic frequency $f=\omega/2\pi$ reduces to $\rho_a = 0.2\ T\ |Z|^2$.

II. The two-dimensional (2-D) Earth

The simple scalar relationship described by (4) breaks down in the

presence of lateral geoelectric inhomogeneities, as is usually the case with the real Earth. Inhomogeneities can be classified as two-dimensional (2-D) or three-dimensional (3-D), depending on the variation of conductivity along the x- and y-axes. The 2-D geoelectric structure is a very important special case and assumes that conductivity is discontinuous along one of the x- or y-axes only. This implies that the inhomogeneities are caused by conductivity variations across elongate interfaces; an approximation with several real life analogues such as faults and dykes. Assuming that the strike of the electrical inhomogeneity lies along the x-axis, (x-derivatives equal to zero), Maxwell's equations decouple into a Transverse Electric (TE) mode with the E-field parallel to the strike of the inhomogeneity and $H_x=0$, and a Transverse Magnetic (TM) mode that requires the E-field to be perpendicular to strike, (H-field parallel), and $E_x=0$. The two modes are popularilly refered to as the H-polarization and the E-polarization respectively. The field behaviour is governed by the equations:

(a) E-polarization:
$$\partial H_z/\partial y - \partial H_y/\partial z = \sigma E_x$$
$$\partial E_x/\partial z = i\omega\mu_0 H_y$$
$$\partial E_x/\partial y = -i\omega\mu_0 H_z$$

(b) H-polarization:
$$\partial E_z/\partial y - \partial E_y/\partial z = i\omega\mu_0 H_x$$
$$\partial H_x/\partial z = \sigma E_y$$
$$\partial H_x/\partial y = -\sigma E_z$$

The H-polarization comprises (H_x, E_y, E_z) field components and corresponds to the poloidal mode of spherical geometries, while the E-polarization comprises (E_x, H_y, H_z) and corresponds to the toroidal mode.

The strike of the 2-D conductivity discontinuity is usually at an arbitrary angle with respect to the directions of measurement of

mutually orthogonal **E** and **H** field components; these are usually taken to be the geographic coordinates x-north, y-east while z is facing down. In this case the curl equations indicate that the EM field components retain their coupled form. Each electric field component will depend on both H_x and H_y magnetic components and the simple scalar relationship (7) will transform into the linear system (e.g. Cantwell, 1960)

$$\begin{bmatrix} E_x \\ E_y \end{bmatrix} = \begin{bmatrix} Z_{xx} & Z_{xy} \\ Z_{yx} & Z_{yy} \end{bmatrix} \begin{bmatrix} H_x \\ H_y \end{bmatrix} \quad <--> \quad \mathbf{E} = \mathbf{Z}\,\mathbf{H} \tag{8}$$

where **Z** is a tensor of rank 2. This relationship will be valid in a source free laterally inhomogeneous medium and will be independent of the source polarization and position (Word et al, 1970). It is possible to transform **Z** so that it corresponds to the decoupled form (7) by rotating the coordinate axes of the experimental frame of reference, until they correspond to the principal directions of the inhomogeneous structure (parallel and perpendicular to strike). Then, the diagonal elements of **Z** will vanish and (8) will read (Word et al, 1970)

$$\begin{bmatrix} E_x \\ E_y \end{bmatrix} = \begin{bmatrix} 0 & Z_{xy} \\ -Z_{yx} & 0 \end{bmatrix} \begin{bmatrix} H_x \\ H_y \end{bmatrix} \tag{9}$$

which is the 2-D analogue of (7). The rotation operator is an orthogonal matrix of the form

$$\mathbf{R} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \tag{10}$$

representing a clockwise rotation. The operation

$$\mathbf{Z}'(\theta) = \mathbf{R}^T\,\mathbf{Z}\,\mathbf{R} \tag{11}$$

rotates the tensor about the z-axis, on the (x,y) plane. Under rotation by π radians, the elements of **Z** trace ellipses on the complex plane, that all have the same size and shape and are defined by the conjugate radii (Word et al, 1970)

$$Z_3 = (Z_{xy} + Z_{yx})/2 \qquad (12a)$$

$$Z_4 = (Z_{xx} - Z_{yy})/2. \qquad (12b)$$

The following quantities are rotationally invariant (Word et al, 1970)

$$Z_1 = (Z_{xy} - Z_{yx})/2 \qquad (12c)$$

$$Z_2 = (Z_{xx} + Z_{yy})/2, \qquad (12d)$$

and define the centroids of the rotational ellipses of the off-diagonal elements ($\pm Z_1$), and diagonal elements (Z_2). The principal radii of the rotation ellipses are $Z'_3(\theta_0)$ and $Z'_4(\theta_0)$; θ_0 can be obtained by maximizing Z_3 and is given by (Word et al, 1970)

$$\tan 4\theta_0 = \frac{2\,\mathrm{Re}[Z_3\,Z_4^*]}{|Z_4|^2 - |Z_3|^2}$$

This has become the conventional method of defining the principal components of the impedance tensor. At any angle θ the tensor elements are given by

$$Z'_{xx}(\theta) = Z_2 + Z_3\sin 2\theta - Z_4\cos 2\theta \qquad (13a)$$

$$Z'_{xy}(\theta) = Z_1 + Z_3\cos 2\theta - Z_4\sin 2\theta \qquad (13b)$$

$$Z'_{yx}(\theta) = -Z_1 + Z_3\cos 2\theta - Z_4\sin 2\theta = -Z'_{yx}(\theta + \pi/2) \qquad (13c)$$

$$Z'_{yy}(\theta) = Z_2 - Z_3\sin 2\theta - Z_4\cos 2\theta = Z'_{yy}(\theta + \pi/2) \qquad (13d)$$

When the tensor is referred to the principal directions, its anti-diagonal elements become

$$Z'_{xy}(\theta_0) = Z_1 + Z_3\cos 2\theta \qquad (13e)$$

$$Z'_{yx}(\theta_0) = -Z_1 + Z_3\cos 2\theta = -Z'_{xy}(\theta + \pi/2) \qquad (13f)$$

It must be appreciated that (11) is a similarity transformation that reduces **Z** to an anti-diagonal form through a series of orthogonal reductions. Therefore at the angle θ_0, i.e. along the principal directions of the conductivity structure, the anti-diagonal elements of $\mathbf{Z}'(\theta_0)$ are the eigenvalues of **Z**, and the operation is analogous to an eigenvalue-eigenvector decomposition. The characteristic vectors of **Z**

are the column vectors of the rotation operator $\mathbf{R}(\theta_0)$ and are real (e.g. see Swift, 1967). This is equivalent to imposing linear polarization on the **E** and **H** fields. Details on the subject will be discussed later.

Two additional parameters can be extracted from $\mathbf{Z}'(\theta_0)$; these are defined by the principal radii of the rotation ellipses, in the complex plane, and indicate the measure of departure of $\mathbf{Z}'(\theta_0)$ from the pure anti- diagonal form of the 2-D impedance. The skew parameter (Swift 1967), is given by

$$s = |Z_2(\theta_0)|/|Z_1(\theta_0)| \qquad (14)$$

and measures the distance from the origin of the diagonal elements' rotation ellipse, normalized by the distance from the origin of the off-diagonal elements' rotation ellipses. For the zero-diagonal **Z'** s=0. The ellipticity parameter is what it says for the rotation ellipse, and is defined by

$$b = |Z_4(\theta_0)|/|Z_3(\theta_0)|. \qquad (15)$$

It follows that b=0 for the 2-D case. The departure of s and b from zero measures the degree of 3-dimensionality and coupling between field components (Word et al, 1970).

III. The three-dimensional (3-D) Earth

This is the most complex case of MT data analysis. With the electrical properties of the Earth varying along all x,y and z directions, the conductivity will assume the tensor form

$$\underline{\sigma} = \begin{bmatrix} \sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ \sigma_{yx} & \sigma_{yy} & \sigma_{yz} \\ \sigma_{zx} & \sigma_{zy} & \sigma_{zz} \end{bmatrix} \qquad (16)$$

Macroscopically, with no derivative vanishing, the curl equations detail detail how the EM field

components couple along the directions of the coordinate system. The vector Helmholz equation (3) can be solved only by separating the contributions of a normal part, (a stack of horizontal uniform layers), and an anomalous part, (embedded 3-D conductivity structure) as in Raiche 1974, Weidelt 1975, Ting and Hohman 1980. The separations

$$\underline{\sigma} = \sigma_n + \sigma_a,\ k^2 = k_n^2 + k_a^2,\quad \mathbf{E} = \mathbf{E}_n + \mathbf{E}_a$$

lead to the formation of two vector Helmholz equations

$$\text{curl.curl}\mathbf{E}_n + k_n^2\,\mathbf{E}_n = -i\omega\mu_0 J_e, \text{ and} \qquad (17a)$$

$$\text{curl.curl}\mathbf{E}_a + k_a^2\,\mathbf{E}_a = -k_n^2\,\mathbf{E} \qquad (17b)$$

The solution of (17a) corresponds to a 1-D conductivity structure. The solution of (17b) often takes the form of a Fredholm integral equation (Raiche 1974, Weidelt 1975, Hohman 1983), in which the dependence of the EM field as a function of the position **r** over the Cartesian space is given by a 3x3 Green's tensor of rank 2. Weidelt (1975) gives the equation

$$\mathbf{E}_a(r_0) = \mathbf{E}_n(r_0) - \int k_a^2(\mathbf{r})\mathbf{G}(r_0|\mathbf{r}).\mathbf{E}(\mathbf{r})dv \qquad (18)$$

where dv denotes a volume element.

The immediate implication of (16), and (18) (and the equivalent expressions for the magnetic field), is that three orthogonal components of **E** and **H** will be present at all times, and fully coupled along the directions of the experimental frame of reference. The linear relationship between the total electric and magnetic field components must assume the form

$$\begin{bmatrix} E_x \\ E_y \\ E_z \end{bmatrix} = \left[\begin{array}{cc|c} Z_{xx} & Z_{xy} & Z_{xz} \\ Z_{yx} & Z_{yy} & Z_{yz} \\ \hline Z_{zx} & Z_{zy} & Z_{zz} \end{array}\right] \begin{bmatrix} H_x \\ H_y \\ H_z \end{bmatrix} \quad <\!\!-\!\!-\!\!> \mathbf{E}^3 = \mathbf{Z}^3\mathbf{H}^3 \qquad (19)$$

The upper left hand side partition of $\mathbf{Z}^3$ refers to the *observable* ($\mathbf{Z}^2$) and the rest to the *undeterminable* parts of the tensor. In (19), the element Z_{zz} denotes the coupling between the vertical field components,

that are generated within the Earth, and has the same meaning with the Z_{xx} tensor element for instance. Relationship (19) is not only a mathematical peculiarity. The vertical component of the electric field has been measured and has been shown to be correlated with the horizontal components (e.g. Bahr, 1983); the existence of the full tensor relationship (19) has thus been confirmed experimentally. The problem with MT data analysis over 3-D inhomogeneities is that while the true relationship between the EM field components is given by (19), only the observable partition is available.

CHAPTER 1

AN ANALYSIS OF THE MAGNETOTELLURIC OPERATORS IN THREE DIMENSIONS

PART 2:

IMPEDANCE TENSOR TRANSFORMATIONS IN MULTIDIMENSIONS

IV. Introduction - The properties of the impedance tensor under conventional rotations.

The conventional rotation operation and resulting parameter set possess several problematic features; the first to recognize and describe them was Eggers (1982). Also, Spitz (1985) gave a rigorous mathematical analysis of the behaviour of $\mathbf{Z}^2$ under conventional rotations, and their inadequacy in the presence of 3-D inhomogeneities. The problems associated with the conventional parameter set can be summarized as follows:

(a) The conventional rotation is insensitive to the trace of $\mathbf{Z}^2$, as can be seen from the absence of Z_2 in the expressions for the anti-diagonal tensor elements in equations (13). This means that the addition of an arbitrary constant to the diagonal elements leaves the estimation of the principal components of $\mathbf{Z}^2$ unaffected, and implies that the extremal values may not be located. It is quite alarming that the apparent resistivities, i.e. the quantities we seek to obtain from the measurements, are subject to such a hazard. Moreover, equations (13) show that the rotation of the diagonal and anti-diagonal elements is independent; $Z'_{xx}(\theta_0)$ depends on the complex scalar Z_2, while $Z'_{xy}(\theta_0)$ on Z_1.

(b) The skew parameter, as defined from equation (14) is insensitive to the phase of Z_2, and is defined only by the magnitude. As Eggers (1982) shows, (a) and (b) indicate that by varying the phase of Z_2 only, there

will be an infinite set of 'impedances' that yield an identical set of parameters according to the conventional definitions.

(c) The conventional ellipticity is insensitive to the sense of rotation (clockwise or counterclockwise) of the loci defining the rotation ellipses; it is defined by the magnitude of Z_4 and Z_3 only, while their phases cannot be taken into account. By reversing the diagonal elements, there exist two 'impedances' that yield identical conventional parameter sets. Thus, in cases of elliptical polarization, we are unable to tell the handedness of the EM field rotation.

From a topological point of view, if each of the $Re(Z_{ij})$ and $Im(Z_{ij})$, i,j=x,y parts of the tensor elements is assigned one degree of freedom (topological dimension), then, $\mathbf{Z}^2$ is defined in a space with as many as eight dimensions over the real field. The conventionally rotated anti-diagonal elements depend on six degrees of freedom (the complex scalars Z_1,Z_3, and Z_4), and the diagonal elements also depend on six degrees of freedom (Z_2,Z_3, and Z_4) out of the required eight. This means that the conventional rotation operation is topologically incomplete. Moreover, since the rotation of the diagonal and anti-diagonal elements is different by Z_1 and Z_2 respectively, their respective parameter space is also different. This is another way of arguing that the conventional operators act independently on the diagonal and anti-diagonal elements of $\mathbf{Z}^2$. The conventional azimuth depends on four degrees of freedom out of the initial eight, and likewise the skew and ellipticity parameters depend on four degrees of freedom each. This gross topological incompletness of the conventional parameter set, gives rise to all the aforementioned problems.

One question that needs to be answered, is why does the analysis work in

the case of the 2-D approximation. This is because of the physical constraints we impose on such a conductivity distribution, that demand the vanishing of the trace (diagonal elements), when the tensor is referred to the principal directions. One will observe that at this unique angle, the conventional azimuth and the principal (anti-diagonal) components, depend on four degrees of freedom, that are identical, i.e. depend on the same parameter space. However, imposition of such conditions (e.g. linear polarization), is possible only under certain conditions that require the conductivity to be discontinuous along vertical planes, and absence of EM field component coupling in a direction perpendicular to strike (i.e. a medium that is uniform perpedicular to strike). Thus, for *any* conductivity distribution that generates elliptically polarized fields along the principal horizontal directions, the eigenvectors of the rotated impedance tensor can no longer be taken to be real. The conventional analysis becomes topologically incomplete and may display some or all of the problematic features discussed above. Elliptical polarization may be produced by 2-D conductivity distributions that are non-uniform perpendicular to strike, and, certainly by all 3-D features. It also means that since the eigenvectors of the impedance tensor are not real, the rotation *cannot* locate the extremal principal values; the diagonal elements of the tensor will therefore not vanish, and the analysis will produce an unreliable parameter set.

IV.1 The relationship between Z^2 and Z^3 under conventional rotations

In this section, I will attempt to describe a physical approach to the problems encountered when rotating the observable tensor Z^2, in the

presence of 3-D structures. The group O(3) is a subset of the O(n) group of nxn orthogonal matrices with n(n-1)/2 independent parameters. For n=2 there is only one independent parameter, namely the angle θ of the rotation matrix **R**. For n=3 there exist three independent parameters so that the group matrices can operate in a space (D^3) with rotations about all three axes. Only group elements with determinant +1 are considered (rotations only, no reflections). This is the subgroup Special Orthogonal SO(3) (e.g. Normand, 1980). Its matrices are explicitly written as

$$\mathbf{R}_x = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{bmatrix} \qquad \mathbf{R}_y = \begin{bmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{bmatrix} \qquad \mathbf{R}_z = \begin{bmatrix} \cos\gamma & -\sin\gamma & 0 \\ \sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Combinations of rotations about the x,y,z axes can be performed by direct multiplications of the SO(3) matrices.

Let us assume for the moment that the full $\mathbf{Z}^3$ impedance tensor is given. The 3-D generalization of the conventional rotation (11) about the z-axis, in a counterclockwise sense, is written as

$$\mathbf{Z}^{3\prime}(\gamma) = \mathbf{R}_z \, \mathbf{Z}^3 \, \mathbf{R}_z^T$$

The assumption that there exists an angle α_0 such that the diagonal elements $Z^{3\prime}_{xx}(\gamma_0)=Z^{3\prime}_{yy}(\gamma_0)=0$, vis.

$$\mathbf{Z}^{3\prime}(\gamma_0) = \begin{bmatrix} 0 & Z_{xy} & Z_{xz} \\ Z_{yx} & 0 & Z_{yz} \\ Z_{zx} & Z_{zy} & Z_{zz} \end{bmatrix}$$

<u>cannot</u> be realized for an actual impedance tensor. The rotation is confined to the (x,y) plane, and because the operator $\mathbf{R}_z$ spans invariant subspaces in the vertical planes, the basic linear MT relationship (19) immediately reveals that the vertical magnetic field component H_z will remain coupled to *both* horizontal field components E_x, E_y. A similar study of the admittance tensor $\mathbf{A}^3$ and the equation $\mathbf{H}^3 = \mathbf{A}^3\mathbf{E}^3$ reveals that the vertical electric field component E_z will always remain coupled to

both horizontal magnetic components H_x, H_y. No vertical field component vanishes at any time, in any direction, because, the total electric and magnetic fields are tilted away from the horizontal, and the tilts have, in general, different orientations in space (e.g Weidelt 1975, Hohman 1983), so that the three-dimensional coupling can be sustained. The *same* features will characterize the observable tensor $\mathbf{Z}^2$. We have to appreciate that we measure coupled field components $[E_x, E_y, H_x, H_y]$, and that this coupling will be retained in the estimated $\mathbf{Z}^2$ irrespective of the analytical procedure implemented for its computation. The trace of $\mathbf{Z}^2$ will therefore not vanish so that there will be *no* real angle, in the horizontal plane, such that $Z^{3\prime}_{xx}(\gamma_0)=Z^{3\prime}_{yy}(\gamma_0)=0$ for a 3-D Earth.

The three dimensional tensor $\mathbf{Z}^3$ belongs to the algebraic class of reversible matrix operators within the physical space D^3, represented by the Cartesian orthonormal basis $\{\mathbf{x},\mathbf{y},\mathbf{z}\}$, (Berdichevsky and Zhdanov, 1984); this basis posseses the topology of the three dimensional ball of unit radius D^3. The observable partition $\mathbf{Z}^2$ belongs to the algebraic class of reversible matrix operators in the Cartesian basis $\{\mathbf{x},\mathbf{y}\}$ embedded in the three-dimensional physical space D^3. With such a definition of the physical space and field properties, we have absolutely no reason to anticipate that the principal directions along which the impedance tensor will anti- diagonalize, will be confined to the (x,y) plane, as is the case with the 2-D approximation. In general, we should expect situations where these directions are *oblique* to the experimental frame of reference. The direction cosines that map such a coordinate system onto the horizontal basis $\{\mathbf{x},\mathbf{y}\}$ will produce non-vanishing diagonal elements in $\mathbf{Z}^2$. The conventional analysis is unable to decouple such contributions because it cannot sense them in the vertical planes. Let us observe that in the

three-space, the dyad $\mathbf{Z}^2$ can be thought of as comprising three sub-tensors of the form

$$\mathbf{Z}_1^2 = \begin{bmatrix} Z_{xx} & Z_{xy} \\ Z_{yx} & Z_{yy} \end{bmatrix} \quad \mathbf{Z}_2^2 = \begin{bmatrix} Z_{xx} & Z_{xz} \\ Z_{zx} & Z_{zz} \end{bmatrix} \quad \mathbf{Z}_3^2 = \begin{bmatrix} Z_{yy} & Z_{yz} \\ Z_{zy} & Z_{zz} \end{bmatrix}$$

each one viewed in the same way as our familiar $\mathbf{Z}^2(=\mathbf{Z}_1^2)$ tensor, i.e. we can decompose $\mathbf{Z}^3$ into 'three 2-D cases', in each of the three orthogonal planes. These subtensors can be rotated, and would possess properties similar to our familiar observable impedance; the principal directions of the rotation ellipses in the vertical planes (x,z) and (y,z) admit the same interpretation as the familiar direction θ_0 in the (x,y) plane. Rotations about any one of the x,y,z axes can be done using the matrices of the SO(3) group.

In conclusion, it appears that a realizable (and realistic) rotation of the 3-D impedance tensor must be such that the vertical and horizontal field components decouple. Thus, we have the following two requirements (a) Any suitable directions, involving real angles, with reference to an orthonormal (Cartesian) frame must be defined in the three-space, and (b) A composite rotation scheme is needed.

Several authors have attempted to define a complete set of parameters for the estimation of the extremal values of $\mathbf{Z}^2$. As Eggers (1982) was the first to recognize and describe the problems, he was also the first to attempt to resolve them, by parameterizing the impedance in terms of an eigenstate formulation, using a modified biorthogonal analysis (Lanzcos, 1961). Spitz (1985), investigated the eigenstate formulation in more detail, but adopted a Cayley factorization approach, while LaTorraca et al (1986) adapt and apply the Singular Value Decomposition (SVD) method to extract the extremal values of $\mathbf{Z}^2$. I single out these

three papers as the most important recent contributions to the theory and understanding of the impedance tensor properties. These analyses, however, do not describe the rotational properties of Z^2 in a three-space, while one of them (the SVD) displays a problematic theoretical basis. Moreover, widely different analytical approaches, are implemented, so that some confusion may be caused in interpreting the resulting parameter sets, especially when one attempts a comparison. This work generalizes the conventional analysis to impedance coordinate transformation (rotation) in three dimensions, by making use of rotation group theory elements. It is hoped that the resulting theory will provide a comprehensive means of understanding the behaviour of the impedance tensor in the presence of 3-D structures and estimating the parameters of the principal directions of the impedance tensor Z^2, henceforth referred to as the intrinsic coordinate system. For the observable tensor Z^2, the composite rotation scheme needed, can be provided by the 2x2 complex matrices of the three-dimensional Special Unitary group SU(2).

V. Realizable coordinate transformations using the group SU(2)

V.1 The Special Unitary Group SU(2)

The subject is very involved and extensive, so as to make the most willing attempt give a generous outline untenable for the purposes of this work. For a very good introduction to the group and its applications , the reader is referred to Arfken (1985). An advanced and detailed study can be found in Normand (1980), while the early work of Wigner (1959), is a very informative reference for the advanced reader;

also, Rose (1957) presents a detailed and readable account of the rotation group. Only absolutely essential information will be provided here, concerning the vector space structure and the topology of the group, as well as its relationship to SO(3). The group SU(2) is a subset of the **U**(n) group of nxn unitary matrices with n^2-1 independent parameters. As with SO(3), the condition det{**U**(n)}=+1 imposes rotations only, no reflections, and defines the Special Unitary (Unimodular) group SU(n). For n=2 there exist 3 independent parameters, as with the SO(3).

The vector space structure of the group is Euclidean. Consider the three anti-commuting Pauli spin matrices

$$\mathbf{s}_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \qquad \mathbf{s}_2 = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} \qquad \mathbf{s}_3 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

obeying the relationships

$$\mathbf{s}_i\mathbf{s}_j + \mathbf{s}_j\mathbf{s}_i = 2\delta_{ij}\mathbf{I}, \qquad \text{(anticommutation)}$$

$$\mathbf{s}_i\,\mathbf{s}_j = i\mathbf{s}_k, \qquad \text{(cyclic permutation of indices)}$$

$$(\mathbf{s}_i)^2 = \mathbf{I}.$$

These matrices are Hermitian, traceless, and linearly independent over the real field. Together with **I** they form a complete set. More importantly, their commutation relationships are such, that they can form a *vector basis* of the 3-D vector space, over the real field. The SU(2) group elements can be generated by the Pauli matrices, through the expression

$$\mathbf{U}(\psi,n^i) = \exp(-i\tfrac{1}{2}\psi\, n^i\, \mathbf{s}_i)$$

where n^i are real numbers such that $\sum_{i=1}^{3}(n^i)^2 = 1$, and ψ is a real parameter. The above equation can also be expressed in terms of an Euclidean inner product as (Normand, 1980)

$$\mathbf{U}(\psi,n^i) = \mathbf{U}(\psi,\mathbf{n};\mathbf{k}) = \exp(-i\tfrac{1}{2}\psi\, \mathbf{n}.\mathbf{s}), \qquad \mathbf{n} = n^i\mathbf{e}_i \qquad \mathbf{n}.\mathbf{s} = n^i\mathbf{s}_i$$

where $\mathbf{k}=(\mathbf{e}_i)$ denotes an orthonormal basis of a three-dimensional

Euclidean vector space. By this definition, the matrix element $\mathbf{U}(\psi,\mathbf{n};\mathbf{k})$ in SU(2), explicitly depends on the number ψn^i, taken for convenience to be the component of a vector in an orthonormal basis. By expanding the exponential in a Maclaurin series, separating even and odd powers and making use of the commutation relationships of the Pauli s's one obtains

$$\mathbf{U}(\psi,n^i) = \mathbf{I}\cos(\tfrac{1}{2}\psi) - in^i s_i \sin(\tfrac{1}{2}\psi)$$

The space of parameters ψ , n^i with a one to one correspondence with $\mathbf{U}(\psi,n^i)$ is the three-dimensional ball D^3 of radius 2π (ψ can be restricted as $[0,2\pi]$, and n^i can take either positive or negative values), so that $\mathbf{U}(\psi,n^i) \rightarrow -\mathbf{U} = (2\pi-\psi,-n^i)$ correspond to two diametrically opposite points within the ball, and $\mathbf{U}^{-1} = \mathbf{U}(\psi,-n^i)$. The group is capable of operating in a 3-D space although it comprises 2x2 matrices.

A second useful parametrization is in terms of the *Cayley-Klein* parameters. Expressing the unitary and unimodular character of the complex 2x2 matrix **U**, one obtains

$$\mathbf{U} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in SU(2) \quad <--> \quad \det\{\mathbf{U}\} = ad - cb = 1$$

$$\bar{\mathbf{U}} = \begin{bmatrix} a^* & c^* \\ b^* & d^* \end{bmatrix} = \mathbf{U}^{-1} = \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

which is equivalent to requiring that two complex numbers $a=x_1+iy_1$, $b=x_2+iy_2$, fulfil the conditions

$$\mathbf{U} = \begin{bmatrix} a & b \\ -b^* & a^* \end{bmatrix}, \qquad |a|^2 + |b|^2 = 1, \qquad \forall\ \mathbf{U} \in SU(2)$$

Thus, there is a one to one correspondence between the elements of SU(2) and the surface points of the four-dimensional sphere of unity radius, denoted by S^3. With this parametrization, two diametrically opposite points correspond to **U** and **-U**.

The parametrization of the group in terms of both the Cayley-Klein parameters and the Pauli spin matrices, defines a topology equivalent to the topology of the four-dimensional sphere S^3, and only locally equivalent to the topology of the three-dimensional ball of unit radius D^3. In contrast, SO(3) possesses the topology of the three-dimensional ball of unit radius D^3 only.

Let $E_3(e_1,e_2,e_3)$ be a right-handed orthonormal basis, such that any two vectors $\mathbf{v}_1$,$\mathbf{v}_2$ spanning the plane (e_1,e_2) obey the relationships

$$\mathbf{v}_1\cdot\mathbf{v}_2 = ||\mathbf{v}_1||.||\mathbf{v}_2||\cos\psi \quad , \qquad \mathbf{v}_1 \times \mathbf{v}_2 = \sin\psi\, ||\mathbf{v}_1||.||\mathbf{v}_2||e_3$$

so that a rotation angle ψ is uniquely defined. In such a base, the correspondence between the groups SO(3) and SU(2) is homomorphic so that

$$\mathbf{R}_1(\alpha) \leftrightarrow \mathbf{U}_1(\alpha,n^1), \qquad \mathbf{R}_2(\beta) \leftrightarrow \mathbf{U}_2(\beta,n^2), \qquad \mathbf{R}_3(\gamma) \leftrightarrow \mathbf{U}_3(\gamma,n^3)$$

In this form, the SU(2) group matrices describe rotations in a counterclockwise sense, like their correspondent SO(3) operators. The establishment of the correspondence is an involved subject, and only a heuristic example will be given here, for the conventional base $E_3(e_1,e_2,e_3)=E_3\{\mathbf{x},\mathbf{y},\mathbf{z}\}$. Take for instance the operator $\mathbf{U}_3$, and carry out a unitary transformation on each of the Pauli $\mathbf{s}$'s to obtain

$$\begin{aligned} \mathbf{U}_3 x\mathbf{s}_1\bar{\mathbf{U}}_3 &= x\cos\gamma\, \mathbf{s}_1 + x\sin\gamma\, \mathbf{s}_2 \\ \mathbf{U}_3 y\mathbf{s}_2\bar{\mathbf{U}}_3 &= -y\sin\gamma\, \mathbf{s}_1 + y\cos\gamma\, \mathbf{s}_2 \\ \mathbf{U}_3 z\mathbf{s}_3\bar{\mathbf{U}}_3 &= z \end{aligned} \qquad (20)$$

Let $\mathbf{T}$ be the zero-trace matrix

$$\mathbf{T} = x\mathbf{s}_1 + y\mathbf{s}_2 + z\mathbf{s}_3 = \begin{bmatrix} z & x-iy \\ x+iy & -z \end{bmatrix} \qquad (21a)$$

under the unitary transformation $\mathbf{T}' = \mathbf{U}_i\, \mathbf{T}\, \bar{\mathbf{U}}_i$. Since the trace is invariant under a unitary transformation, any matrix $\mathbf{T}'$ in the new coordinate system must have the form

$$\mathbf{T}' = x'\mathbf{s}_1 + y'\mathbf{s}_2 + z'\mathbf{s}_3 = \begin{bmatrix} z' & x'-iy' \\ x'+iy' & -z' \end{bmatrix} \qquad (21b)$$

From (20) and (21) one may establish that

$$x' = x\cos\tfrac{1}{2}\gamma - y\sin\tfrac{1}{2}\gamma$$

$$y' = x\sin\tfrac{1}{2}\gamma + y\cos\tfrac{1}{2}\gamma$$

$$z' = z$$

i.e., the unitary transformation $\mathbf{U}_3(\tfrac{1}{2}\gamma)$ is equivalent to the rotation operator $\mathbf{R}_z(\gamma)$. The full correspondence for this basis is therefore

$$\mathbf{R}_x \leftrightarrow \mathbf{U}_x, \qquad \mathbf{R}_y \leftrightarrow \mathbf{U}_y \qquad \mathbf{R}_z \leftrightarrow \mathbf{U}_z,$$

with

$$\mathbf{U}_x = \begin{bmatrix} \cos\tfrac{1}{2}\alpha & -i\sin\tfrac{1}{2}\alpha \\ -i\sin\tfrac{1}{2}\alpha & \cos\tfrac{1}{2}\alpha \end{bmatrix} \quad \mathbf{U}_y = \begin{bmatrix} \cos\tfrac{1}{2}\beta & -\sin\tfrac{1}{2}\beta \\ \sin\tfrac{1}{2}\beta & \cos\tfrac{1}{2}\beta \end{bmatrix} \quad \mathbf{U}_z = \begin{bmatrix} \exp(-i\tfrac{1}{2}\gamma) & 0 \\ 0 & \exp(i\tfrac{1}{2}\gamma) \end{bmatrix}$$

The half angle notation simply means that whatever transformation is effected by SO(3) through an angle ψ, is equivalently effected by SU(2) through an angle $\psi/2$. This also forms the basis of the homomorphism between the two groups. For example, as γ in $\mathbf{R}_z$ varies from 0 to 2π , $\gamma/2$ in $\mathbf{U}_z$ varies from 0 to π, so that

$$\mathbf{R}_z(\gamma+2\pi) = \mathbf{R}_z(\gamma), \text{ and } \mathbf{U}_z(\tfrac{1}{2}\gamma + \pi) = -\mathbf{U}_z(\tfrac{1}{2}\gamma).$$

Both $\mathbf{U}_z(\gamma/2)$ and $-\mathbf{U}_z(\gamma/2)$ correspond to $\mathbf{R}_z(\gamma)$, i.e. the representation is not unique. This will have implications in the implementation of rotations on $\mathbf{Z}^2$ that will be discussed in due time. What is important however, is that the establishment of this correspondence means that the known representations of SU(2) automatically provide us with the representations of SO(3). In the ensuing discussion the half angle notation will be dropped for simplicity, nonetheless it must not be forgotten.

<u>The SU(2) group multiplication law:</u> It is apparent that as with SO(3), any product of the SU(2) group elements is a member of the group, and describes composite rotations about the axes of an orthonormal Euclidean

(Cartesian) frame. Let us go back to the exponential representation of SU(2), and closely observe the vector basis. Then, we find that this is a special (3-D) case of n^2-1 dimensional vector space formed by the set S(n), of the nxn complex Hermitian and traceless matrices. This set forms a complete group with a Lie algebra structure, and a Euclidean vector space structure, both closely related to the group structure of SU(n) (e.g. Normand, 1980). Next, notice that studying exponentially represented groups is equivalent to studying the products of their exponential representations. Denote by $\mathbf{S}_j = \frac{1}{2}\,\mathbf{s}_j \in S(2)$ any component vector in the basis E_3. Then, introduce the Baker-Campbell-Hausdorf (BCH) formula (e.g. Normand, 1980), which gives an explicit formal expression of the matrix **c**, in terms of the matrices **a** and **b**, such that exp(i**a**) exp(i**b**) = exp(i**c**). Since any element **U** in SU(2) can be expressed in the form $\mathbf{U} = \exp(-i\mathbf{S}_j)$, one may obtain the group multiplication law

$$\mathbf{U}_i\,\mathbf{U}_j = \exp(-i\mathbf{S}_i)\exp(-i\mathbf{S}_j) = \mathbf{U}_k = \exp(-i\mathbf{S}_k), \qquad \mathbf{S}_k \in S(2).$$

The BCH formula expresses that for any infinitesimal rotation by $\mathbf{S}_i$ and $\mathbf{S}_j$, the elements of $\mathbf{S}_k$ can be expressed in terms of the repeated commutators of $\mathbf{S}_i$ and $\mathbf{S}_j$. Normand (1980) gives the first terms of this expansion as

$$\mathbf{S}_k = \mathbf{S}_i + \mathbf{S}_j - i[\mathbf{S}_i, \mathbf{S}_j] - \frac{1}{12}([\mathbf{S}_i,[\mathbf{S}_i,\mathbf{S}_j]] + [\mathbf{S}_j,[\mathbf{S}_j,\mathbf{S}_i]])\ldots$$

where the symbol [..] stands for a commutation relationship. It follows that the group multiplication law is formally defined if one knows the commutation rules of any matrices $\mathbf{S}_i$ and $\mathbf{S}_j$ in S(2), which requires knowledge of the Lie algebra of S(2). This can be shown to be isomorphic to that of SU(2). The basic commutation relationships in the vector basis S(2) are defined by $-i[\mathbf{S}_i, \mathbf{S}_j] = \varepsilon_{ijk}\,\mathbf{S}_k$, with ε_{ijk}, the Levi- Civita symbol for the totally antisymmetric

tensor of order three. Naturally, these commutation rules are a consequence of the commutation rules of the Pauli s matrices. By using the above expansion we see that

$$S_k = S_i + S_j + \varepsilon_{ijk} S_k + O(S_k) \qquad (22)$$

where $O(S_k)$ represents the less significant contributions to the rotation of the vector basis. The important point of the BCH formula concerning SU(2) applications, is that infinitesimal rotations using any product $U_i U_j$ are equivalent to second order infinitesimal rotations in the *entire* space spanned by the group, since they involve the whole basis S(2). It will be seen that this form of the group multiplication law is very important in defining the SU(2) operators, suitable to process the impedance tensor Z^2.

The vector spaces spanned by SU(2) will also be of paramount importance in the ensuing discussion, and need to be introduced here, in some detail. It has already been indicated that SU(2) possesses Cartesian and spherical topologies. It can be shown that group can be represented in a Cartesian and a Spherical vector basis, because these spaces are isomorphic. Introduce the two-dimensional complex orthonormal (biorthogonal) basis

$$\underline{\varepsilon}_{\pm} = 2^{-1/2}(\mathbf{x} + i\mathbf{y}) \qquad (23)$$

with properties $\underline{\varepsilon}^{*}_{\pm} \underline{\varepsilon}_{\pm} = 1$, and $\underline{\varepsilon}^{*}_{\pm} \underline{\varepsilon}_{\mp} = 0$. The use of this basis in Electromagnetism is described by (Jackson 1975), and was also used by Eggers (1982). The unit vectors in $\underline{\varepsilon}_{\pm}$ represent circularly polarized waves of opposite handedness. For any two-dimensional Euclidean subspace $\{E_2\} \subset \{E_n\}$ spanned by the orthonormal vectors $\underline{\varepsilon}_{\pm}$, one can define the vectors

$$\underline{\varepsilon}_1 = 2^{-1/2}(\underline{\varepsilon}_{+} + \underline{\varepsilon}_{-}), \quad \text{and} \quad \underline{\varepsilon}_2 = -i(2^{-1/2})\,(\underline{\varepsilon}_{+} - \underline{\varepsilon}_{-})$$

that have only real components, and form an orthonormal basis $\underline{\varepsilon}_j$ in the

space $\{E_2\}$. This is isomorphic, and therefore can be identified with the basis E_2. Notably, in this particular basis only, the diagonal exponential matrix

$$\mathbf{U}(\psi) = \begin{bmatrix} \exp(-i\psi) & 0 \\ 0 & \exp(i\psi) \end{bmatrix} \text{ reduces to the form } \mathbf{R} = \begin{bmatrix} \cos\psi & -\sin\psi \\ \sin\psi & \cos\psi \end{bmatrix}$$

The isomorphic character of the two bases can be consolidated in the reversible, and therefore one-to-one unitary transformation

$$2^{-\frac{1}{2}} \begin{bmatrix} x + iy \\ x - iy \\ 2^{\frac{1}{2}} z \end{bmatrix} = 2^{-\frac{1}{2}} \begin{bmatrix} 1 & i & 0 \\ 1 & -i & 0 \\ 0 & 0 & 2^{\frac{1}{2}} \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} \qquad (24a)$$

which has the important property of preserving the properties of the z-axis. Next, consider that the reversible unitary transformation

$$\begin{bmatrix} 2^{-\frac{1}{2}}(x + iy) \\ z \\ 2^{-\frac{1}{2}}(x - iy) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 2^{-\frac{1}{2}}(x + iy) \\ 2^{-\frac{1}{2}}(x - iy) \\ z \end{bmatrix} \qquad (24b)$$

facilitates, and completes a transition from a Cartesian to a *spherical* basis, in which the expression

$$\tau_{1m}(\phi,\theta) = r^{-1} \begin{cases} 2^{-\frac{1}{2}}(x + iy) & m = 1 \\ z & m = 0 \\ 2^{-\frac{1}{2}}(x - iy) & m = -1 \end{cases}$$

Is defined to be the normalized harmonic polynomial *(solid harmonic)* of order n=1 (e.g. Normand 1980). The solid harmonic is related to the spherical harmonic of order n=1 as

$$Y_{1m} = \sqrt{(3/4\pi)}\ r^{-1}\ \tau_{1m}$$

(e.g. Rose, 1957), and the factor $\sqrt{(3/4\pi)}\ r^{-1}$ is invariant under rotations. The Condon-Shortley phase factor of $(-1)^m$ is not included since it is unnecessary for our purposes. Y_{1m} defines points on the surface of the unit sphere, and obeys the orthogonality condition

$$Y^*_{1m}(\phi,\theta) = Y_{1,-m}(\phi,\theta), \text{ and } \qquad Y_{nm}\, Y^*_{nm} = \delta_{nm}.$$

It must be obvious by now, that the isomorphism of the spaces and the homomorphism of the SO(3) and SU(2) groups suggests that Cartesian and

Spherical transformations and tensor algebras are closely related. A formal description of this relationship can be found in Normand (1980, pp. 213ff). It is shown that from any cartesian tensor **T**, in the orthonormal space $E_3(e_1,e_2,e_3)$, one can define a mapping **T′** of E_3 into the set of 2x2 complex matrices. The simplest form in which the tensor **T′** can be visualised is that of equations (21). By using the standard relationships between the spherical harmonics and the harmonic polynomials (e.g. Normand p. 463) and assuming that r=1, we can write **T′** as

$$\sqrt{4\pi/3}\ \mathbf{T}' = 2^{-\frac{1}{2}}\begin{bmatrix} 2^{\frac{1}{2}}\cos\theta & \sin\theta\exp(-i\phi) \\ \sin\theta\exp(i\phi) & 2^{\frac{1}{2}}(-\cos\theta) \end{bmatrix} \qquad (25)$$

The spherical harmonics are in effect spherical tensors; they are related to the the elements of the tensor **T′** as indicated by (24b), and thence to the elements of **T**, as indicated by (24a). They also constitute a representation of SO(3), in the vector space $L^2(S^2,d_2\mathbf{n})$ of square integrable complex functions in S^2 , with S^2 denoting the closed ball of radius one. Thus, it is concluded that **T′** constitutes a representation of SU(2). This means that rotations about the axes of a Cartesian frame are isomorphic to, and therefore can be described by, the spherical harmonic of the first order. It is noteworthy (and important) that the spherical vector space is natural to the MT problem; the vector Helmholz equations, in general, admit solutions in spherical coordinates.

In the basis (23), or in the spherical coordinate basis, one defines the component vectors

$$\mathbf{s}_+ = \mathbf{s}_1 + i\mathbf{s}_2 = \begin{bmatrix} 0 & 2 \\ 0 & 0 \end{bmatrix}, \quad \text{and} \quad \mathbf{s}_- = \mathbf{s}_1 - i\mathbf{s}_2 = \begin{bmatrix} 0 & 0 \\ 2 & 0 \end{bmatrix}$$

that obey the commutation relationships

$$[s_+ , s_-] = 4s_3 , \qquad [s_3, s_\pm] = \pm 2s_\pm$$

so that

$$S_+ = \tfrac{1}{2}s_+ , \quad S_- = \tfrac{1}{2}s_- , \text{ and } S_3 = \tfrac{1}{2}s_3 ,$$

obeying the commutation relationships

$$[S_+, S_-] = S_3 , \qquad [S_3, S_\pm] = \pm S_\pm ,$$

form the structure constants of a Lie algebra, in the spherical vector basis, that is isomorphic, and therefore can be represented by the Lie algebra of SU(2).

V.2 A Decomposition of the impedance tensor Z^2 using the group SU(2)

V.2.1 On the choice of a suitable rotation operator

We are given a Cartesian tensor Z^2 with z-projections, where by the term z-projections I define the influences (coupling) of the vertical and horizontal field components, measured in the horizontal basis $E_2\{x,y\}$. We are asked to transform the tensor in a 3-D space, so that its z-projections maximize simultaneously, in two orthogonal directions. This will be achieved through a rotation of the z-axis about one of the x- or y-axes, simultaneously with a rotation of the basis $E_2\{x,y\}$, carrying the measurements, about the z-axis. The objective requires a minimum of two Euler rotations, namely a precession (rotation about the z-axis), followed by a nutation, which comprises a rotation about the y-axis. Let us, however, consider the vector basis. In section V.1, the orthonormal right-handed basis $E_3\{e_1,e_2,e_3\}$ has been identified with the right-handed basis $E_3\{x,y,z\}$ so that

$$x \rightarrow n^1 s_1 \qquad y \rightarrow n^2 s_2 \qquad z \rightarrow n^3 s_3$$

This is the most widely used convention in Quantum Mechanics, although

different approaches exist in literature (e.g. Murnaghan 1938, Wigner 1959). In this basis, the required Euler rotations will assume the form

$$\mathbf{U}_c = \mathbf{U}_z\,\mathbf{U}_y = \begin{bmatrix} \exp(-i\gamma) & 0 \\ 0 & \exp(i\gamma) \end{bmatrix} \begin{bmatrix} \cos\beta & -\sin\beta \\ \sin\beta & \cos\beta \end{bmatrix}$$

The operator $\mathbf{U}_c$ is inconsistent with the spatial properties of $\mathbf{Z}^2$ (and $\mathbf{A}^2$), because $\mathbf{U}_z$ operates on the complex orthonormal basis $\underline{\varepsilon}_{\pm}=2^{-\frac{1}{2}}(\mathbf{x}+i\mathbf{y})$, while $\mathbf{Z}^2$ and $\mathbf{A}^2$ are obtained in the real, linearly polarized cartesian basis $E_2\{\mathbf{x},\mathbf{y}\}$. As a consequence, the operator $\mathbf{U}_c$ is unsuitable for their treatment.

The easiest way arround this problem is to choose a more suitable vector basis. Because we are constrained by the real property of the x- and y-axes, and because no measurements are taken along the vertical direction, the obvious choice is to let the z-axis be an imaginary coordinate. Here, I follow the system of Murnaghan (1938), with the difference that I require the vector basis to be right-handed. The new basis is defined by a cyclic permutation of the indices of the conventional basis, as

$$n^1\mathbf{s}_1 = n^1\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \qquad n^2\mathbf{s}_2 = n^2\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \qquad n^3\mathbf{s}_3 = n^3\begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}$$

in the 3-D complex space

$$\mathbf{T}'' = n^1\mathbf{s}_1 + n^2\mathbf{s}_2 + n^3\mathbf{s}_3 = \begin{bmatrix} x & y-iz \\ y+iz & -x \end{bmatrix}$$

One shows (with the method of section V.1) that in the new basis, the required rotation operators become

$$\mathbf{U}_z = \begin{bmatrix} \cos\gamma & -\sin\gamma \\ \sin\gamma & \cos\gamma \end{bmatrix} \qquad \mathbf{U}_y = \begin{bmatrix} \cos\beta & -i\sin\beta \\ -i\sin\beta & \cos\beta \end{bmatrix}$$

so that $\mathbf{U}_z$ is real, and suitable to rotate matrices obtained in the real, linearly polarized basis $E_2\{\mathbf{x},\mathbf{y}\}$. Finally, for reasons explained

in section V.6.2, I will rather make use of the form

$$\mathbf{U} = \mathbf{U}_z \, \bar{\mathbf{U}}_y = \begin{bmatrix} \cos\gamma & -\sin\gamma \\ \sin\gamma & \cos\gamma \end{bmatrix} \begin{bmatrix} \cos\beta & i\sin\beta \\ i\sin\beta & \cos\beta \end{bmatrix} \qquad (26)$$

with the overbar denoting the transposed conjugate matrix.

V.2.2 On the application of the rotation operator

A rotation of the type $\mathbf{Z}'(\beta,\gamma) = \bar{\mathbf{U}}(\beta,\gamma)\ \mathbf{Z}\ \mathbf{U}(\beta,\gamma)$ will not reduce the tensor **Z** to an antidiagonal form (the dimensionality superscript will henceforth be dropped). The necessary and sufficient condition that a complex matrix **X** can be diagonalized (or anti-diagonalized, depending on the physical conditions), by a unitary operator **U**, is that **X** be *normal*, i.e. commute with its transposed conjugate (e.g. Normand, 1980), in the sense $\mathbf{X}\,\bar{\mathbf{X}} = \bar{\mathbf{X}}\,\mathbf{X}$. However, we know that **Z** is not a normal matrix.

Being unable to rotate **Z** directly, I will circumnavigate the problem by seeking to rotate ansatz normal matrices that can be constructed from **Z**. However, such a rotation is at least difficult, because it requires the simultaneous variation of two parameters, i.e. the angles β and γ. A promising way to overcome this problem, is to introduce the operation as a decomposition of **Z** into a product

$$\mathbf{Z} = \mathbf{U}(\beta_1,\gamma_1)\ \mathbf{M}\ \bar{\mathbf{U}}(\beta_2,\gamma_2) \qquad (27)$$

such that **M** is *antidiagonal* and contains the non-increasing characteristic values of **Z**, and $\mathbf{U}(\beta_1,\gamma_1)$, $\mathbf{U}(\beta_2,\gamma_2)$ are unitary matrices of the form (26). In order to justify such a decomposition I shall make use of the following

Proposition 1: Let **X** be a regular nxn complex matrix of rank n. Then there exists one nxn unitary matrix **U**, a nxn unitary matrix **V**, and a nxn antidiagonal complex matrix **S** such that

$$\mathbf{S} = \bar{\mathbf{U}}\,\mathbf{X}\,\mathbf{V} \qquad \text{and} \qquad \mathbf{X} = \mathbf{U}\,\mathbf{S}\,\bar{\mathbf{V}}$$

if and only if the products $\mathbf{X}\,\bar{\mathbf{X}}$ and $\bar{\mathbf{X}}\,\mathbf{X}$ are normal matrices. The successive antidiagonal entries of **S** are non-increasing.

Proof: The matrix $\mathbf{X}\,\bar{\mathbf{X}}$ has an eigenvalue-eigenvector decomposition

$$\mathbf{X}\,\bar{\mathbf{X}} = \mathbf{U}\,\mathbf{D}\,\bar{\mathbf{U}}$$

where the nxn matrix **U** is unitary and the matrix **D** diagonal with positive entries, if and only if it is normal. The matrix **D** is defined to be the product $\mathbf{S}\,\bar{\mathbf{S}}$, with $(\mathbf{D})_{ii} > (\mathbf{D})_{jj}$, i>j. Let $\mathbf{Y}=\mathbf{S}^{-1}$. Then, it can be easily verified that $\mathbf{Y}\,\mathbf{D}\,\bar{\mathbf{Y}} = \mathbf{I}_n$. Define the nxn matrix $\bar{\mathbf{V}} = \mathbf{Y}\,\bar{\mathbf{U}}\,\mathbf{X}$, so that $\mathbf{V} = \bar{\mathbf{X}}\,\mathbf{U}\,\bar{\mathbf{Y}}$, and

$$\bar{\mathbf{V}}\,\mathbf{V} = \mathbf{Y}\,\bar{\mathbf{U}}\,\mathbf{X}\,\bar{\mathbf{X}}\,\mathbf{U}\,\bar{\mathbf{Y}} = \mathbf{Y}\,\mathbf{D}\,\bar{\mathbf{Y}} = \mathbf{I}_n$$

which shows that **V** is unitary. Then, again from the definitions

$$\mathbf{U}\,\mathbf{S}\,\bar{\mathbf{V}} = \mathbf{U}\,\mathbf{S}\,\mathbf{Y}\,\bar{\mathbf{U}}\,\mathbf{X} = \mathbf{U}\,\bar{\mathbf{U}}\,\mathbf{X} = \mathbf{X}$$

because **U** is unitary. Thus, it is shown that the decomposition is realizable. Note, however, that owing to the fact that **S** is anti-diagonal, $\mathbf{S}\,\bar{\mathbf{S}} \neq \bar{\mathbf{S}}\,\mathbf{S}$. Thus, the converse must be shown, i.e. the case $\bar{\mathbf{X}}\,\mathbf{X}$. Again, this matrix admits the eigenvalue-eigenvector decomposition

$$\bar{\mathbf{X}}\,\mathbf{X} = \mathbf{V}\,\mathbf{D}'\,\bar{\mathbf{V}}$$

with the nxn matrix **V** unitary, if and only if it be normal. **D′** is defined by the product $\bar{\mathbf{S}}\,\mathbf{S}$ with $(\mathbf{D}')_{ii} < (\mathbf{D}')_{jj}$, i>j. Then, $\bar{\mathbf{Y}}\,\mathbf{D}'\,\mathbf{Y} = \mathbf{I}_n$, and consequently, we have

$$\bar{\mathbf{U}}\,\mathbf{U} = \bar{\mathbf{Y}}\,\bar{\mathbf{V}}\,\bar{\mathbf{X}}\,\mathbf{X}\,\mathbf{V}\,\mathbf{Y} = \bar{\mathbf{V}}\,\mathbf{D}'\,\mathbf{Y} = \mathbf{I}_n$$

which shows that **U** is unitary. As before, define the product

$$\mathbf{U}\,\mathbf{S}\,\bar{\mathbf{V}} = \mathbf{X}\,\mathbf{V}\,\mathbf{Y}\;\;\mathbf{Y}\,\bar{\mathbf{V}} = \mathbf{X}\,\mathbf{V}\,\bar{\mathbf{V}} = \mathbf{X},$$

which shows that the decomposition is realizable. It is interesting to

note that for an anti-diagonal characteristic matrix **S**, the eigenvalues of $\mathbf{X}\bar{\mathbf{X}}$ are non-increasing, while the eigenvalues of $\bar{\mathbf{X}}\mathbf{X}$ are increasing. This anti-symmetrical behaviour can be observed to hold in both legs of the proof and has profound implications in the analysis of the impedance tensor because it introduces constraints in the behaviour of a physical matrix with antidiagonal eigenvalues. The same philosophy as above can be used to prove the following

Proposition 2: Let $\mathbf{X}'$ be a regular nxn complex matrix of rank n. Then, there exists one nxn unitary matrix $\mathbf{U}'$, one nxn unitary matrix $\mathbf{V}'$, and one *diagonal* matrix $\mathbf{S}'$ such that

$$\mathbf{X}' = \mathbf{U}' \, \mathbf{S}' \, \bar{\mathbf{V}}'$$

if and only if $\bar{\mathbf{X}}'\mathbf{X}'$ is a normal matrix. The successive diagonal entries of $\mathbf{S}'$ will be non-increasing. Such a decomposition possesses all the properties of the Singular Value Decomposition, extended to a complex matrix, and unitary operators. In the case of the SVD, the anti-symmetric properties revealed in the preceding proof vanish.

We see that it is possible to factor a nxn complex matrix in two ways. In the case of any 2x2 regular complex matrix, it can easily be verified that both the products $\mathbf{Z}\bar{\mathbf{Z}}$ and $\bar{\mathbf{Z}}\mathbf{Z}$ are Hermitian matrices, and therefore normal by definition. Thus the impedance tensor **Z** can be factored in both ways presented above. I shall concentrate on the study of the properties of decomposition (27), and, once enough background is accumulated, I will show that this is the only physically realizable alternative.

By virtue of Proposition 1, one can postulate that the matrices $\mathbf{C}_1 = \mathbf{Z}\bar{\mathbf{Z}}$ and $\mathbf{C}_2 = \bar{\mathbf{Z}}\mathbf{Z}$ admit an eigenvalue-eigenvector decomposition of the form

$$\mathbf{C}_1 = \mathbf{U}(\beta_1,\gamma_1)\ \mathbf{D}_1\ \bar{\mathbf{U}}(\beta_1,\gamma_1), \text{ and} \qquad (28a)$$

$$\mathbf{C}_2 = \mathbf{U}(\beta_2,\gamma_2)\ \mathbf{D}_2\ \bar{\mathbf{U}}(\beta_2,\gamma_2) \qquad (28b)$$

while (27) gives

$$\mathbf{M} = \bar{\mathbf{U}}(\beta_1,\gamma_1)\ \mathbf{Z}\ \mathbf{U}(\beta_2,\gamma_2) \qquad (28c)$$

The diagonal matrices $\mathbf{D}_1$ and $\mathbf{D}_2$ result by rotation of $\mathbf{C}_1$ and $\mathbf{C}_2$ through angles (β_1,γ_1) and (β_2,γ_2) respectively. The diagonalization of $\mathbf{C}_1$ and $\mathbf{C}_2$ leaves the Euclidean norms $||\mathbf{Z}\bar{\mathbf{Z}}||= ||\mathbf{C}_1||$ and $||\bar{\mathbf{Z}}\mathbf{Z}||= ||\mathbf{C}_2||$ invariant. Consider the characteristic value matrix

$$\mathbf{M} = \begin{bmatrix} 0 & \mu_1 \\ \mu_2 & 0 \end{bmatrix}, \quad |\mu_1| > |\mu_2|, \quad \mu_j\mu_j^* = r_j^2.$$

Then,

$$\mathbf{D}_1 = \begin{bmatrix} r_1^2 & 0 \\ 0 & r_2^2 \end{bmatrix} \quad (29a), \qquad \mathbf{D}_2 = \begin{bmatrix} r_2^2 & 0 \\ 0 & r_1^2 \end{bmatrix} \quad (29b)$$

The vector lengths $||\mathbf{Z}\bar{\mathbf{Z}}||$ and $||\bar{\mathbf{Z}}\mathbf{Z}||$ are equal since $\text{Trace}[\mathbf{D}_1]=\text{Trace}[\mathbf{D}_2] = \text{Trace}[\mathbf{C}_1] = \text{Trace}[\mathbf{C}_2]$ (a unitary transformation leaves the trace and the determinant invariant). Then the length of $\mathbf{Z}$ can be understood to be

$$||\mathbf{D}_1^{1/2}|| = ||\mathbf{D}_2^{1/2}|| = (r_1^2 + r_2^2)^{1/2}$$

$\mathbf{C}_1$ and $\mathbf{C}_2$ are the ansatz matrices constructed from $\mathbf{Z}$ that can be diagonalized using unitary operators. These operators are combined in (27) or (28c) with the characteristic values of $\mathbf{Z}$. Thus $\mathbf{U}(\beta_1,\gamma_1)$ and $\mathbf{U}(\gamma_2,\beta_2)$ provide two sets of directions that communicate through the characteristic value matrix of $\mathbf{Z}$. The structures of the matrices $\mathbf{C}$ are mutually exclusive, i.e. $\mathbf{C}_1$ and $\mathbf{C}_2$ do not commute; these matrices can be viewed as transforms of the impedance tensor associated with distinct states of the EM field.

In order to make things more clear, observe that the product $\mathbf{Z}\bar{\mathbf{Z}}$ is similar to the product of a complex scalar with its complex conjugate, $cc^*=g$; c^* represents an image (reflection) of c with respect to the

real axis. In the case of the 2x2 complex matrix **Z**, the 'image' is represented by $\bar{\mathbf{Z}}$. However, in contrast to the simple mirror image of Im(c) in the two-dimensional geometry of the two-parameter complex plane, **Z** comprises four complex numbers, and should be defined in an eight-parameter hyperplane, with each one of the $Re(Z_{ij})$, $Im(Z_{ij})$, i,j = x,y assigned one degree of freedom; thus, the parameter space has as many as eight dimensions over the real field. The vector space of the transposed conjugate matrix $\bar{\mathbf{Z}}$ shares its topology with that of **Z**, but is totaly anti-symmetric. We shall call this space the *dual* to **Z**, and write the associated dual orthonormal vector basis, symbolically, as $E^{*3}(e^{*1},e^{*2},e^{*3})$; clearly, (e^{*i}) is a contravariant set to (e_i). The formation of the product $\mathbf{C}_1 = \mathbf{Z}\bar{\mathbf{Z}}$ can be viewed as a mapping of **Z** into the Hermitian vector space of square integrable functions $L^2(S^2,d_2\mathbf{n})$ (namely a Hilbert space) with reduced dimensionality; in this space, $\mathbf{C}_1$ can properly be treated with the spherical tensor (solid harmonic) operators. The opposite of course is true, in what concerns the product $\mathbf{C}_2$. Thus the whole procedure is a self-selection process that reduces the dimensionality of the parameter space by 'cancelation' of symmetries through orthogonal operations (reduction of the number of degrees of freedom). The products $\mathbf{C}_1$ and $\mathbf{C}_2$ are 2x2 Hermitian matrices of the form

$$\mathbf{C} = \begin{bmatrix} k & l+im \\ l-im & n \end{bmatrix} \qquad \text{with k,l,m,n real}$$

i.e. they depend on four degrees of freedom only. Moreover, since $[\mathbf{C}_1, \mathbf{C}_2] \neq 0$, they comprise single states, with four degrees of freedom. In terms of the Cayley-Klein parametrization, the matrices of SU(2) also depend on four degrees of freedom. Thus, there is a

one-to-one correspondence in the topology of **C** and **U** (the four dimensional open ball S^3), and the parameters in (28a,b) form complete sets. Next, observe that the matrices $\mathbf{U}(\beta_1,\gamma_1)$ and $\mathbf{U}(\beta_2,\gamma_2)$ both define eight degrees of freedom, so that their correspondence with **Z** should also form a complete set. It has been convenient to introduce (28) as a decomposition, but one should always remember that it also can be a function of continuously varying parameters (the rotation angles). It can easily be verified that for arbitrary values of the rotation angles, each element of the tensor $\mathbf{Z}' = \bar{\mathbf{U}}(\beta_1',\gamma_1')\ \mathbf{Z}\ \mathbf{U}(\beta_2',\gamma_2')$ will always be given as a linearly independent combination of eight real parameters. The product, defines 32 independent real numbers; the combination $\mathbf{X} = [\mathbf{Z}\ \mathbf{U}(\beta_2',\gamma_2')]$ defines eight independent complex products, while there exist 16 independent complex products in the combination $\mathbf{Z}' = \bar{\mathbf{U}}(\beta_1',\gamma_1')\ \mathbf{X}$. Each element of $\mathbf{Z}'$ is given by a linear combination of four complex (eight real) parameters, so that the rotation will depend on a full complement of eight degrees of freedom. Moreover, the Cayley-Klein parameterization provides a symmetry that can be used to verify that the rotation of the diagonal and antidiagonal elements is not independent.

V.3 On the interpretation of the rotation angles.

The eigenvectors of the **C** matrices are the *column* vectors of **U**, and one can write $\mathbf{U} = [\mathbf{V}^+\ \mathbf{V}^-]$. Because $\bar{\mathbf{U}}\mathbf{U} = \mathbf{I}$, the eigenvectors are orthogonal, so that $\bar{\mathbf{V}}^{\pm}\ \mathbf{V}^{\pm} = 1$, and $\bar{\mathbf{V}}^{\pm}\ \mathbf{V}^{\mp} = 0$. The left eigenvector $\mathbf{V}^+$, from (26b), can be written as

$$\mathbf{V}^+ = \begin{bmatrix} a \\ -b^* \end{bmatrix} = \begin{bmatrix} \cos\gamma & -\sin\gamma \\ \sin\gamma & \cos\gamma \end{bmatrix} \begin{bmatrix} \cos\beta \\ i\sin\beta \end{bmatrix} = \mathbf{U}_z \begin{bmatrix} \cos\beta \\ i\sin\beta \end{bmatrix} \qquad (30)$$

Likewise, the right eigenvector **V** can be written as

$$\mathbf{V}^{-} = \begin{bmatrix} b \\ a^{*} \end{bmatrix} = \begin{bmatrix} -\sin\gamma & \cos\gamma \\ \cos\gamma & \sin\gamma \end{bmatrix} \begin{bmatrix} \cos\beta \\ i\sin\beta \end{bmatrix} = [U_z \; s_1] \begin{bmatrix} \cos\beta \\ i\sin\beta \end{bmatrix} \qquad (31)$$

The (original) matrix s_1 is a reflection operator, that interchanges the coordinate axes (produces $\pi/2$ shift in the origin of the rotation and inversion of its handedness). This implies that any two components of the two-component eigenvectors exist in mutually orthogonal directions. Two pairs of orthogonal directions are defined by $\{(\beta, \gamma), (\beta, \gamma + \pi/2)\}$, that comprise two independent intrinsic coordinate systems for **Z**. One important point, that must be appreciated and can clearly be seen in (30) and (31), is that the eigenvectors $\mathbf{V}^{\pm}$ are in themselves rotation operators for two-component orthogonal vectors.

V.3.1 The definition of the rotation parameter γ.

In interpreting the rotation angles β and γ, it is convenient to think in Spherical frames of reference. Thus, it becomes clear that at least one of the β , γ angles must assume the role of an azimuthal direction, while the other will be defined as a longitudinal (polar) dependence. By definition of the rotation operator (26), the azimuthal direction should be identified with the angle γ. In order to study it in more detail, introduce the biorthogonal basis $\underline{\varepsilon}_{\pm}$, defined by equation (23).

The isomorphism between the vector bases $\underline{\varepsilon}_{\pm}$ and E_2, as observed in (24), allows one to define the eigenvectors in either one. Let the cartesian frame be the linearly polarized basis $E_2\{\mathbf{x},\mathbf{y}\}$; in order to understand

the behaviour of the eigenvectors in this basis, consider a product of the form

$$|x \;\; y|\, U = |x \;\; y|\, [V^{+} \; 0] + |x \;\; y|\, [0 \;\; V^{-}].$$

Then, the left eigenvector V^{+}, for instance, gives

$$\begin{aligned} V^{+}(\beta,\gamma:E_2) &= \mathrm{Re}(V^{+}) + i\,\mathrm{Im}(V^{+}) \\ \mathrm{Re}(V^{+}) &= (\cos\gamma\, x + \sin\gamma\, y)\cos\beta \qquad (32) \\ \mathrm{Im}(V^{+}) &= (-\sin\gamma x + \cos\gamma\, y)\sin\beta \end{aligned}$$

The real and imaginary parts of V^{+} in (32) describe the orthonormal linearly polarized basis E_2 rotated by an angle γ, while the whole eigenvector is rotated by an angle β. In the isomorphic basis (23), the eigenvectors are expressed as

$$V = V_{+}\,\underline{\varepsilon}_{+} + V_{-}\,\underline{\varepsilon}_{-}, \qquad (33)$$

where $V_{+} = V^{+}\underline{\varepsilon}^{*}_{+}$ and $V_{-} = V^{-}\underline{\varepsilon}^{*}_{-}$ are circularly polarized components in $\underline{\varepsilon}_{\pm}$. These vectors can readily be put in the form

$$\begin{aligned} V_{+} &= [\cos\beta\, \exp(-i\gamma) \qquad \sin\beta\, \exp(i\gamma)], \\ V_{-} &= [\cos\beta\, \exp(-i[\gamma+\pi/2]) \qquad \sin\beta\, \exp(i[\gamma+\pi/2])] \end{aligned}$$

reminiscent of the spherical coordinate system; this form can be thought of, as an intermediate stage during the transition from the Cartesian to the Spherical systems. Thus, by (32) and (33), γ is readily defined as an azimuthal angle in the horizontal plane.

V.3.2 The definition of the rotation parameter β.

I now proceed to investigate the nature of the angle β as a physical direction. Let us assume that two principal directions (β_0, γ_0) are given. Then, one can rotate the experimental frame of reference on the (x,y) plane by an angle γ_0, so that $(x,y,z) \rightarrow (x_1,y_1,z)$. Now consider that from the point of view of an observer in the frame

(x_1,y_1,z) the γ angles vanish; in the new coordinate frame of reference $\gamma = 0$. Then, the eigenvectors of the type (32) reduce to the form

$$\mathbf{V}^{+} = \mathbf{x}\cos\beta + i\mathbf{y}\sin\beta \ , \quad \text{and} \quad \mathbf{V}^{-} = \mathbf{y}\cos\beta + i\,\mathbf{x}\sin\beta$$

with $\bar{\mathbf{V}}^{\pm}\,\mathbf{V}^{\mp} = 0$. The length of the unit vectors **x** and **y** remains unchanged, because the horizontal transformation is isometric, i.e. maps the vector space onto itself. However, the 'reduced' eigenvectors are also given in terms of isometric *projections* of unit vectors. The question is projections of what, and from where, since we agree that β is not a physical angle (azimuth) in the horizontal plane. The answer is provided by considering the spherical coordinate system, and the solid harmonics as rotation operators. Then, if γ is an azimuth, β will assume the role of a longitudinal (polar) dependence in the spherical system. This implies that β can formally be defined as a physical angle in the vertical planes of the Cartesian frame (x_1,y_1,z). In order to overcome the objection that no vertical field measurements are taken, consider that since the coordinate axes **y** and **z** can be images of each other (equation 24b), in isomorphic albeit different coordinate systems, one may anticipate that if quantities defined in the vertical and horizontal axes interact, then, information components associated with the vertical coordinates can be conveyed as information components in the horizontal axes. These information components will assume the form of *projections* of the z-axis onto the horizontal plane; the opposite of course is true, by symmetry.

Thus, the answer I propose is schematically given in Figure 1, where three Cartesian frames are depicted. Axes (x,z) belong to the experimental frame of reference; the y-axis is not shown for the sake

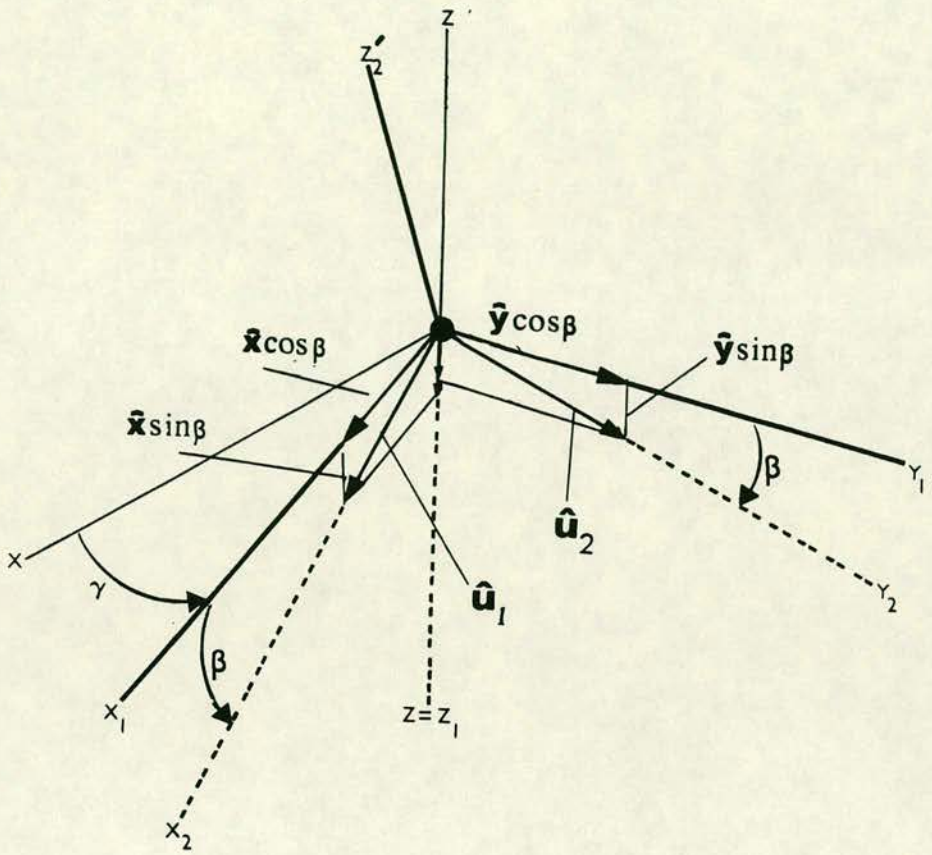


FIGURE 1 Diagram depicting the result of the rotation operator U_{zy}. Axes (x,z) belong to the experimental Cartesian frame of reference C; the y-axis is not shown for the sake of clarity. Axes $(x_1,y_1,z_1=z)$ belong to the Cartesian frame C_1, resulting by a horizontal rotation of C through an angle γ. Axes (x_2,y_2) correspond to the Cartesian frame C_2, dipping at an angle β with respect to C_1; the z_2-axis lies on the direct line of sight of the z-axis, and thus cannot be seen. Frame C can be thought of as resulting by independent rotations about x_1 and y_1. Thus, the z_2'-axis belongs to the frame (x_2,y_1,z_2'), resulting by a rotation about y_1 .

of clarity in the diagram. $(x_1,y_1,z_1=z)$ refer to the horizontally rotated frame, in which $\gamma = 0$. Finally, axes (x_2,y_2,z_2) belong to an orthonormal frame resting at an angle β with respect to the horizontal plane (x_1,y_1). Let $\mathbf{u}_1$ and $\mathbf{u}_2$ be two orthogonal unit vectors in (x_2,y_2,z_2), so that $\mathbf{u}_1.\mathbf{u}_2 = 0$. Then, because $|\mathbf{u}_1| = |\mathbf{u}_2| = |\mathbf{x}| = |\mathbf{y}|$, their projections onto the axes of the (x_1,y_1,z) frame, are the components of the reduced eigenvectors. Thus, the eigenvectors $\mathbf{V}^{\pm}$ are the sum of one horizontal component $\mathbf{u}_i\cos\beta$, and one vertical component $\mathbf{u}_j\sin\gamma$, j=1,2, where $\mathbf{u}_j$ belong to a non-horizontal intrinsic coordinate system of **Z**. The eigenvectors, and, in particular, the vertical component are, given as a combination of projections of two orthogonal unit vectors. This can be understood in terms of the vector basis (23), or the solid harmonics, that couple the two horizontal Cartesian axes. In physical terms, one must note that the orthogonal EM field components are related through antisymmetric rotations by $\pi/2$, vis.

$$E_x \underset{A_{yx}}{\overset{Z_{xy}}{<=\!=\!=>}} H_y \ , \qquad H_x \underset{A_{xy}}{\overset{Z_{yx}}{<=\!=\!=>}} E_y .$$

The rotation by the antisymmetric tensor components will couple their z-projections onto each other. More information is provided in the ensuing discussion.

In order to return to the analysis of **Z**, I will need to refer back to the mode of induction over a 2-D and a 3-D conductivity geometry, and the definition of the three-dimensional impedance tensor $\mathbf{Z}^3$ (also see section IV.2). My basic argument will therefore be that three-dimensional information is included in the impedance tensor when the relationships

$$\begin{aligned} E_x &= Z_{xx} H_x + Z_{xy} H_y + Z_{xz} H_z, \qquad & H_y &= A_{xx} E_x + A_{xy} E_y + A_{xz} E_z \\ E_y &= Z_{yx} H_x + Z_{yy} H_y + Z_{yz} H_z, \qquad & H_x &= A_{yx} E_x + A_{yy} E_y + A_{yz} E_z \end{aligned} \tag{34}$$

are effective due to the 3-D conductivity structure. In (34) we observe

that the vertical field components (H_z, E_z) couple to both horizontal **E** and **H** components. Next, consider the behaviour of the impedance tensor, in the case of the 2-D approximation, when it is referred to the principal directions; the equivalent of (34) then becomes

$$\begin{aligned} &\text{TE mode : } E_x = Z_{xy} H_y + Z_{xz} H_z, \qquad H_y = A_{yx} E_x \\ &\text{TM mode : } E_y = Z_{yx} H_x \qquad\qquad H_x = A_{xy} E_y + A_{zx} E_z \end{aligned} \tag{35}$$

It is obvious that the (H_z, E_z) components do not couple into (E_y, H_y). This is exactly the property that differentiates the two cases. As explained in IV.2 *any* measurements we obtain refer to coupled field components; in a 3-D conductivity geometry, this means *both* horizontal field components of either **E** or **H**, *at any azimuthal direction*. As a consequence, the estimated **Z** will retain this coupling, and should convey information about it. Then, according to the above model, **y**sinβ is the amount of coupling experienced by (E_x, H_x), and **x**sinβ the same for (E_y, H_y), *at the principal azimuthal direction* γ_0. The amount of coupling will be proportional to the magnitude of β, which is equivalent to saying that it is proportional to the magnitude of the vertical field components. The case $\beta = \pi/4$ implies a contribution of the vertical component equal to that of the horizontal, and would result in circularly polarized fields. Now, consider that for any two orthogonal field components $\mathbf{F} = [F_x \quad F_y]$, the eigenvectors $\mathbf{V}^{\pm}$ amount to a rotation operator. Thus, the product $\mathbf{F}.\mathbf{V}^{\pm}$, when referred to the coordinate frame (x_1,y_1,z) will assume the form

$$\mathbf{F}(\gamma_0).\mathbf{V}^{+} = \mathbf{x}\, F_x(\gamma_0)\cos\beta + i\mathbf{y}\, F_y(\gamma_0)\sin\beta ,$$

and

$$\mathbf{F}(\gamma_0).\mathbf{V}^{-} = \mathbf{y}\, F_y(\gamma_0)\cos\beta + i\mathbf{x}\, F_x(\gamma_0)\sin\beta .$$

This means that at the principal azimuthal direction, the contribution of the vertical fields is only due to coupling in the orthogonal component $\mathbf{j}.F_j(\gamma+\pi/2)$, j=x,y. In view of both (34) and (35) one

concludes that the component $y\sin\beta$ will imply the amount of distortion (deviation from the TE or TM mode) in (E_x, H_x) due to the vertical field components (E_z, H_z) coupling with (E_y, H_y), and $x\sin\beta$ will imply the amount of distortion in (E_y, H_y), due to the coupled components (E_x, H_x). When the conductivity distribution becomes 2-D, the EM field separates into two modes; (E_y, H_y) are free of coupling with the vertical fields and independent of (E_x, H_x). Therefore, their contribution in (E_x, H_x) vanishes and $y\sin\beta \to 0$. The independence of the orthogonal field components also implies that $x\sin\beta \to 0$. The eigenvectors $\mathbf{V}^{\pm}$ become pure unit vectors in the Cartesian frame (x_1,y_1,z) which implies linear polarization. It follows immediately, that in the 3-D case, when the frame (x_1,y_1,z) is rotated so as to coincide with (x_2,y_2,z_2), $\beta \to 0$ so that the eigenvectors $\mathbf{V}^{\pm}$ become pure unit vectors in this frame. This implies linearly polarized fields. Therefore, one concludes that if the principal directions (β_0, γ_0) define a non-horizontal intrinsic coordinate frame for $\mathbf{Z}$, then the EM field components in this frame are linearly polarized plane waves.

The results discussed above for the particular frame of reference (x_1,y_1,z) afford a generalization in the experimental frame of reference. To see this one can simply modify (32) using the notation

$$\mathbf{x}_1 = \mathbf{x}\cos\gamma + \mathbf{y}\sin\gamma \,, \qquad \text{and } \mathbf{y}_1 = \mathbf{x} - \sin\gamma + \mathbf{y}\cos\gamma \,,$$

the transformation therefore becoming

$$\mathrm{Re}(\mathbf{V}^{+}) = \mathbf{x}' = \mathbf{x}_1 \cos\beta$$

$$\mathrm{Im}(\mathbf{V}^{+}) = \mathbf{y}' = \mathbf{y}_1 \sin\beta$$

Similarly, the right ($\mathbf{V}^{-}$) eigenvector becomes

$$\mathrm{Re}(\mathbf{V}^{-}) = \mathbf{y}'' = \mathbf{y}_1 \cos\beta$$

$$\mathrm{Im}(\mathbf{V}^{-}) = \mathbf{x}'' = \mathbf{x}_1 \sin\beta$$

If two orthogonal linearly polarized field components $\mathbf{u}_1.F_1 = F_1\cos(\omega t+\phi_1)$, $\mathbf{u}_2.F_2 = F_2\cos(\omega t+\phi_2)$ are given, these can be expressed in the (x,y) plane, in terms of the eigenvectors, as

$$F_1' = \mathbf{x}' + \mathbf{y}' = (\mathbf{u}_1.F_1)\cos\beta + (\mathbf{u}_2.F_2)\sin\beta$$

and

$$F_2' = \mathbf{y}'' + \mathbf{x}'' = (\mathbf{u}_2.F_2)\cos\beta + (\mathbf{u}_1.F_1)\sin\beta\ .$$

One can rotate the vector $\mathbf{F} = [F_1' \quad F_2']$ to coincide with the experimental frame of reference, by $\mathbf{F} = R(-\gamma_0)\ \mathbf{F}'$. Upon use of the trigonometric identity cos(a+b) = cosa cosb - sina sinb, and after some algebra, one obtains (for example) the component

$$\begin{aligned} F_x = \cos(\omega t)[&(x\ (\mathbf{u}_1F_1)\cos\phi_1\cos\gamma + y\ (\mathbf{u}_2F_2)\cos\phi_2\sin\gamma)\cos\beta \\ &+ (x\ (\mathbf{u}_1F_1)\cos\phi_1\sin\gamma + y\ (\mathbf{u}_2F_2)\cos\phi_2\cos\gamma)\sin\beta\ ] \\ -\sin(\omega t)[&(x\ (\mathbf{u}_1F_1)\sin\phi_1\cos\gamma + y\ (\mathbf{u}_2F_2)\sin\phi_2\sin\gamma)\cos\beta \\ &+ (x\ (\mathbf{u}_1F_1)\sin\phi_1\sin\gamma + y\ (\mathbf{u}_2F_2)\sin\phi_2\cos\gamma)\sin\beta\ ] \end{aligned}$$

which expresses the elliptical polarization experienced by the measurement of the orthogonal field components F_j, j=x,y. The field component F_x maximizes at time t=0, and this corresponds to the major axis of its polarization ellipse; it minimizes at $t=\pi/2\omega$, which corresponds to the minor axis. This form of vectors is identical to the concept introduced by Eggers (1982).

If the rotation angle β is kept constant, the variation of the azimuthal angle γ, defines the space of the normalized rotation operator as an ellipse with a major axis equal to $\cos\beta$, and a minor axis equal to $\sin\beta$. The ratio of the minor to the major axis (ellipticity) is given by

$$e = \tan\beta$$

if β is constrained as $-\pi/4 < \beta < \pi/4$. It will soon be seen (Section VI) that β is costrained as above, due to the symmetry implied by the half angle definition of SU(2) and the rotational

symmetry of **Z**. The sign of β determines the sense of rotation of the eigenvectors. Thus, from (32) in can readily be verified that $\beta < 0$ implies a clockwise sense of rotation, while $\beta > 0$ implies a counterclockwise sense. I will argue for the case of the dual interpretation of β as an ellipticity and a physical direction, by proposing that any angle resulting by rotation, must have a physical meaning, if one accepts that mathematics actually describe the physical world. Moreover, if one accepts that the MT source fields are purely (linearly) polarized plane waves, any observed elliptical polarization is generated within the Earth. Ellipticity, as a physical quantity measures the effect and not the cause, which is the geoelectric structure. Therefore, one must conclude that any angle that can be used as a measure of the effect (ellipticity) should also be interpretable in terms of the cause.

The most important novel feature of this analysis is that when elliptically polarized fields, or 3-D conductivity structures are present, the principal directions that antidiagonalize the impedance tensor *are not* confined in the horizontal plane, but dip at an angle β when referred to the principal azimuthal direction γ_0. Furthermore, in view of the above discussion, it is very important to note that from the information provided by the observable impedance tensor **Z** only, (i.e. with no a priori information about the conductivity structure), it is impossible to infer whether the angles β refer to 'real' or 'virtual' directions. By 'virtual directions' I imply orientations in the three-space, along which the EM field separates into two distinct linearly polarized states. These orientations may or may not signify the direction of a geoelectric conductivity boundary, however, this information cannot be extracted from **Z** only. For example, consider the

case of a horizontal rectangular prism embedded in an electrically isotropic matrix. The coupling of the EM field components within the Earth, principally because of the so called 'edge effects', will produce both a vertical impedance element and the β angles, implying the existence of non-horizontal characteristic states. Nonetheless, while the azimuthal γ angles, can be associated with principal directions of the conductivity structure, the angles β have no such counterpart. They constitute 'virtual directions'. In this example case, they are frequency dependent and their magnitude is a function of location with respect to the inhomogeneity; they represent a measure of the distortion of the EM field and the coupling between the vertical and horizontal components. Therefore, the relationship between the β angles and the direction of a conductivity boundary, requires information additional to that provided by the impedance tensor. Their very existence however, and their variation as a function of frequency and location is of paramount importance in assessing the vertical and lateral extent of the inhomogeneity producing them.

V.4 The nature of the intrinsic coordinate system and the characteristic states of the impedance tensor Z.

For reasons that will become apparent immediately, I shall call the intrinsic coordinate frames defined by (28), the E-frame and the H-frame respectively. Upon using the notation $\mathbf{U}(\beta_1,\gamma_1) = \mathbf{U}_E$ and $\mathbf{U}(\beta_2,\gamma_2) = \mathbf{U}_H$ the decomposition (28) can be summarized as

$$\mathbf{Z}\,\bar{\mathbf{Z}} = \mathbf{C}_E = \mathbf{U}_E\,\mathbf{D}_E\,\bar{\mathbf{U}}_E \qquad (36a)$$

$$\bar{\mathbf{Z}}\,\mathbf{Z} = \mathbf{C}_H = \mathbf{U}_H\,\mathbf{D}_H\,\bar{\mathbf{U}}_H \qquad (36b)$$

$$\mathbf{M} = \bar{\mathbf{U}}_E\,\mathbf{Z}\,\mathbf{U}_H \qquad (36c)$$

Equations (36) will henceforth be referred to as the Unitary Decomposition (UD) of the impedance tensor **Z**, and the characteristic values μ_j as the eigenvalues of **Z**. In order to investigate the nature of the two coordinate frames, introduce the antidiagonal matrix

$$\mathbf{L} = \mathbf{M}^{-1} = \begin{bmatrix} 0 & (\mu_2)^{-1} \\ (\mu_1)^{-1} & 0 \end{bmatrix}$$

so that $\mathbf{L}\,\mathbf{M} = \mathbf{M}\,\mathbf{L} = \mathbf{I}$. **L** corresponds to the decomposition

$$\mathbf{L} = \bar{\mathbf{U}}_H(\beta_H, \gamma_H)\ \mathbf{A}\ \mathbf{U}_E(\beta_E, \gamma_E) \qquad (37)$$

of the admittance tensor **A**. Then, note that (29a) results from (28a); by rotation through angles (β_E, γ_E) we locate the direction that maximizes $\mathbf{C}_E$, parallel to its maximum eigenvalue. Likewise, (29b) results from (28b); by rotation through (β_H, γ_H), $\mathbf{C}_H$ minimizes parallel to this direction. Observe however that

$$\bar{\mathbf{M}}\,\mathbf{M} = (\mathbf{L}\,\bar{\mathbf{L}})^{-1} = \begin{bmatrix} (r_2^2)^{-1} & 0 \\ 0 & (r_1^2)^{-1} \end{bmatrix}, \qquad r_1^2 > r_2^2$$

which indicates that this rotation is equivalent to the rotation of $\mathbf{A}\bar{\mathbf{A}}$ by (β_H, γ_H), so that its maximum eigenvalue is located parallel to this direction. In full, the relationships

$$\mathbf{D}_E = \mathbf{M}\,\bar{\mathbf{M}},\quad \mathbf{D}_E^{-1} = \bar{\mathbf{L}}\,\mathbf{L},\quad \mathbf{D}_H = \bar{\mathbf{M}}\,\mathbf{M},\quad \mathbf{D}_H^{-1} = \mathbf{L}\,\bar{\mathbf{L}} \qquad (38a)$$

and

$$\mathbf{D}_E = \mathbf{U}_E\,(\bar{\mathbf{A}}\mathbf{A})^{-1}\,\bar{\mathbf{U}}_E\ ,\ \mathbf{Z}\bar{\mathbf{Z}} = (\bar{\mathbf{A}}\mathbf{A})^{-1} \qquad (38b)$$

$$\mathbf{D}_H = \mathbf{U}_H\,(\mathbf{A}\bar{\mathbf{A}})^{-1}\,\bar{\mathbf{U}}_H\ ,\ \bar{\mathbf{Z}}\mathbf{Z} = (\mathbf{A}\bar{\mathbf{A}})^{-1} \qquad (38c)$$

establish the fact that the impedance and admittance tensors extremize at identical coordinate frames, and, most importantly, the admittance tensor minimizes along the same directions that maximize the impedance. The reciprocal behaviour of the two tensors under the UD, is consistent with their nature. The impedance tensor maps the **H** field onto the **E**

field and describes how its energy is being dissipated; therefore, it must maximize along the 'most resistive' path,and vice versa. Conversely, the admittance tensor maps the **E** field onto the **H** field, and also describes how its energy is lost; therefore, it must minimize along the least conducting (equivalently more resistive) path and vice versa. This information is made available by the UD. An important consequence of the two tensors sharing their intrinsic coordinate systems reciprocally, is that either one must contain information about the other, and, that their characteristic states must be simple reciprocals of each other.

With such an understanding of the behaviour of the UD, consider the usual induction relationship, which by virtue of (36) can be writen as

$$\mathbf{E} = \mathbf{Z}\,\mathbf{H} \xrightarrow{(36)} \mathbf{E} = \mathbf{U}_E\,\mathbf{M}\,\bar{\mathbf{U}}_H\,\mathbf{H},$$

which gives $\bar{\mathbf{U}}_E\,\mathbf{E} = \mathbf{M}\,\bar{\mathbf{U}}_H\,\mathbf{H}$. This can be expanded to give

$$\bar{\mathbf{V}}_E^{+}\,\mathbf{E} = \mu_1\,\bar{\mathbf{V}}_H^{-}\,\mathbf{H} \qquad (39a)$$

$$\bar{\mathbf{V}}_E^{-}\,\mathbf{E} = \mu_2\,\bar{\mathbf{V}}_H^{+}\,\mathbf{H} \qquad (39b)$$

Next, consider the reciprocal induction problem

$$\mathbf{H} = \mathbf{A}\,\mathbf{E} \xrightarrow{(37)} \mathbf{H} = \mathbf{U}_H\,\mathbf{L}\,\bar{\mathbf{U}}_E\,\mathbf{E},$$

which can be developed similarly to give

$$\bar{\mathbf{V}}_H^{+}\,\mathbf{H} = (\mu_2)^{-1}\,\bar{\mathbf{V}}_E^{-}\,\mathbf{E} \qquad (40a)$$

$$\bar{\mathbf{V}}_H^{-}\,\mathbf{H} = (\mu_1)^{-1}\,\bar{\mathbf{V}}_E^{+}\,\mathbf{E} \qquad (40b)$$

Because the eigenvectors $\mathbf{V}^{+}$ and $\mathbf{V}^{-}$ are azimuthally perpendicular, equations (39b), (40a) imply that an **H** field vector, rotated by $\bar{\mathbf{V}}_H^{+}$ to the direction (β_H, γ_H), is mapped onto an **E** field vector rotated by $\bar{\mathbf{V}}_E^{-}$ to the direction$(\beta_E, \gamma_E+\pi/2)$, so that they couple along the 'most conducting' path inside the 3-D Earth. This will be called the *minimum* (-) characteristic state of **Z**. Likewise, Equations (39a), (40b) imply

that an **H** field vector rotated by $\bar{V}_H^-$ to $(\beta_H, \gamma_H+\pi/2)$, is mapped onto an **E** field vector rotated by $\bar{V}_E^+$ to (β_E, γ_E), so that they couple along the 'least conducting' path inside the medium. This corresponds to the *maximum* (+) characteristic state of **Z**. Because, in general, $\gamma_E \neq \gamma_H$, the electric and magnetic field components defined by the characteristic states are *not* perpendicular. The mappings can be summarized as follows

$$\begin{bmatrix} E(\beta_E, \gamma_E) \\ E(\beta_E, \gamma_E + \pi/2) \end{bmatrix} = \begin{bmatrix} 0 & \mu_1 \\ \mu_2 & 0 \end{bmatrix} \begin{bmatrix} H(\beta_H, \gamma_H) \\ H(\beta_H, \gamma_H + \pi/2) \end{bmatrix} \qquad (41)$$

The directions (β_E, γ_E) and (β_H, γ_H) define planes containing orthogonal electric and magnetic field components respectively, with the additional feature that they represent purely (linearly) polarized plane waves, propagating into the Earth medium. The necessary and sufficient condition for such plane waves to couple in the sense described by (41) is

$$\mathbf{E}(\beta_E, \gamma_E) \times \mathbf{H}(\beta_H, \gamma_H) > 0$$

which yields a Poynting vector at the direction of the energy flux of the EM wave propagating into the Earth. In addition, we see that

$$\mathbf{E}(\beta_E, \gamma_E) \cdot \mathbf{H}(\beta_H, \gamma_H) = 0$$

which is a necessary and sufficient condition for the separation of the EM field into distinct, linearly polarized modes.

V.5 The 2-D approximation.

The most important details have already been discussed in V.3, and only some additional physical information is provided here. The 2-D geometry requires the separation of the EM field into TE and TM modes, so that all the transverse components of the field will be perpendicular

everywhere. In the conventional intrinsic frame of reference, this implies that the characteristic states are (a) horizontal, and (b) orthogonal, conditions that require the β angles to vanish. Combine the two modes in a single expression by making use of the Z^3 matrix definition to obtain

$$\begin{bmatrix} E_x \\ E_y \\ E_z \end{bmatrix} = \begin{bmatrix} 0 & Z_{xy} & Z_{xz} \\ -Z_{yx} & 0 & 0 \\ -Z_{yz} & 0 & 0 \end{bmatrix} \begin{bmatrix} H_x \\ H_y \\ H_z \end{bmatrix}$$

The lower left partition of Z^3 defines a null subspace on the (y,z) plane, perpendicular to strike, with the physical interpretation that the E-field is discontinuous (TE mode), or the magnetic field is discontinuous (TM mode) across the conductivity interface. When, along the principal directions $\beta \to 0$, $\mathbf{U}_y \to \mathbf{I}$ so that

$$\mathbf{U}_{zy} = \begin{bmatrix} \cos\gamma & -\sin\gamma \\ \sin\gamma & \cos\gamma \end{bmatrix}$$

i.e. reduces to the form of the orthogonal matrix employed in the conventional analysis. Because the transverse field components must be perpendicular, the TE and TM coordinate frames coincide so that $\gamma_E = \gamma_H$. Then, (36) can be written as

$$\mathbf{M} = \bar{\mathbf{U}}_E(0,\gamma_E)\ \mathbf{Z}\ \mathbf{U}_H(0,\gamma_E)$$

This is equivalent, to the conventional formulation (11), and implies that in the limiting case, the unitary operator becomes indistinguishable from the orthogonal operator, because the physical conditions demand it. The perpendicularity requirement imposes the linear polarization along the principal directions of the infinitely long conductivity discontinuity, since the characteristic vectors $\mathbf{V}^{\pm}$ become real.

V.6 An evaluation of the Singular Value Decomposition (SVD) of the impedance tensor Z.

The application of the SVD to the extremization and analysis of the impedance tensor has been considered by LaTorraca et al (1986). This section is a reappraisal of the physical applicability of the concept to the analysis of MT data.

V.6.1 The concept of the SVD and the impedance tensor.

Consider now the alternative mathematical possibility, the SVD, and write

$$\mathbf{M}' = \bar{\mathbf{U}}'_E \, \mathbf{Z} \, \mathbf{U}'_H \, , \qquad \mathbf{M}' = \text{diag}\,(\mu'_1, \mu'_2) \qquad (42)$$

and the matrices $\mathbf{U}'_E$ and $\mathbf{U}'_H$ unitary. One early distressing feature of (42) is that it cannot reduce to the conventional formulation (11) when the structure becomes 2-D; with $\beta \to 0$, the SVD formulation suggests coupling between parallel EM field components, a consequence that cannot be considered seriously. This is an important loss of generality indicating that the SVD is 'applicable' only in 3-D situations, and unable to perform otherwise, which is at least peculiar. Next consider the inversion

$$(\mathbf{M}')^{-1} = \mathbf{L}' = \bar{\mathbf{U}}'_H \, \mathbf{A} \, \mathbf{U}'_E \qquad (43)$$

in which

$$\mathbf{L}' = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} = \begin{bmatrix} (\mu'_1)^{-1} & 0 \\ 0 & (\mu'_2)^{-1} \end{bmatrix}$$

However, because $|\mu'_1| > |\mu'_2|$ we necessarilly obtain $|\lambda_1| < |\lambda_2|$. i.e. $\mathbf{L}'$ is not the non-increasing singular value matrix of the admittance. (43) *does not* represent the SVD of the admittance tensor. By similar arguments as in (26) and (36) through (38), we reach the conclusion that

under the SVD formulation, the admittance tensor minimizes along the *same* directions that *minimize* the impedance tensor, and vice versa. I question this result as awkward because it contradicts the basic requirement for reciprocality between the two quantities. I suggest that this, plus the fact that the SVD cannot reduce to the conventional formalism illustrate that (42) is unwarranted as a physical decomposition of **Z**. On the other hand, it has been shown that the inversion of (36) provides the true anti-diagonal eigenvalue matrix of **A**. In V.3 I discussed the fundamentally reciprocal and homogeneous behaviour of the two tensors, that can only be conserved if their components are anti-diagonal. This constitutes a physical constraint that *must* be preserved in any physically realizable manipulations of the Earth response operators. However, the SVD and the UD are not totally independent. As it will soon be discussed, they simply operate in different, albeit related spaces; of these two, only the UD space is physical.

V.6.2 The SVD analysis of LaTorraca et al (1986)

The reason I undertake this extensive evaluation of the SVD, is not only to show its inapplicability to the MT tensor analysis, but also to discuss the basis and development of the analysis of LaTorraca et al (1986). The UD and the SVD are inherently similar in their matrix manipulations, so that if identical unitary operators are used, the SVD may lead to the correct results, under certain conditions.

The authors begin their analysis by stating that the SVD results in two two-component normalized and orthogonal **H** vectors $\mathbf{h}_i$ and $\mathbf{h}_j$, and two

two-component normalized and orthogonal **E** vectors $\mathbf{e}_i$ and $\mathbf{e}_j$, with $\mathbf{h}_i\mathbf{h}_j = \delta_{ij}$ and $\mathbf{e}_i\mathbf{e}_j = \delta_{ij}$, such that

$$\mathbf{Z}\,\mathbf{h}_j = p_j\mathbf{e}_j, \qquad (44a)$$

$$\bar{\mathbf{Z}}\,\mathbf{e}_j = p_j\mathbf{h}_j, \qquad p_1 \geq p_2 \geq 0 \qquad (44b)$$

The two equations can be combined to provide

$$\mathbf{Z} = \mathbf{U}'_E\,\mathbf{P}\,\bar{\mathbf{U}}'_H \qquad (44c)$$

where the unitary matrices **U′** are constructed as $\mathbf{U}'_E = [\mathbf{e}_1\ \mathbf{e}_2]$ and $\mathbf{U}'_H = [\mathbf{h}_1\ \mathbf{h}_2]$. Then, they proceed to *adapt* the equations (44) that imply a real singular value matrix, to the form (42), with a complex singular value matrix. The adaptation proceeds by a series of assertions about the form and properties of $\mathbf{e}_j$ and $\mathbf{h}_j$. These entail a counting of parameters in the left and right hand sides of (44c). The authors find ten real parameters on the LHS, i.e. four in each of the unitary matrices **U′**, and two in **P**, while only eight exist in **Z**; they assert that this reflects the fact that (44a,b) remain satisfied if both vectors $(\mathbf{e}_j, \mathbf{h}_j)$ are multiplied by the same phase factor. Considering the redundancy awkward, they propose to even the parameter count by imposing the conditions

$$\mathbf{h}_j.\mathbf{h}_j\ , \qquad \mathbf{e}_j.\mathbf{e}_j \qquad \text{positive definite} \qquad (45a)$$

These conditions can be satisfied by multiplication by the appropriate phase factors (the singular values p_j) that now become complex. According to the authors, this evens the parameters on both sides of (44c), because "the elements in the diagonal matrix, which are the multipliers become complex", and, "there are two parameters each in the restricted **U′**, while there are four parameters in the complex diagonal matrix". Since the vectors $\mathbf{e}_j$, $\mathbf{h}_j$ must be normalized complex Fourier Transforms of real time-varying vectors, condition (45a) imposes on them the structure (32). The unitary matrices **U′** thus become identical to

the rotation operators $\mathbf{U}_z\bar{\mathbf{U}}_y$. However, the orthogonality between, say, $(\mathbf{h}_1, \mathbf{h}_2)$ is satisfied according to the authors, if they are defined as

$$\mathbf{h}_1 = (\beta'_H, \gamma'_H), \qquad \mathbf{h}_2 = (-\beta'_H, \gamma'_H + \pi/2) \tag{45b}$$

This happens, because they consider β' to be a pure ellipticity parameter, and naturally, orthogonality is maintained if the sense of rotation changes in the two eigenvectors. Finally, the half planes for the γ'_E and γ'_H angles are selected by requiring that

$$\mathbf{e}_j \times \mathbf{h}_j > 0 \tag{45c}$$

which yields a Poynting vector at the direction of energy flux of the EM field into the Earth. (45c) is satisfied if

$$-\pi/2 < \gamma'_H < \pi/2, \text{ and } \gamma'_E \leq \gamma'_H < \gamma'_E + \pi \tag{45d}$$

It is quite apparent, I think, that the unitary operators $\mathbf{U}'$ are constructed empirically. Nowhere in their paper do the authors appear to realize that their matrices belong to the SU(2) rotation group, nor do they acknowledge the necessary and sufficient condition of normality, for a complex SVD to be mathematically possible. Therefore they attempt to obtain it by construction, and not analytically. I also find several uneasy features within their analysis. Firstly, I believe it is not correct to assess that "the SVD results in two two-component orthogonal and normalized **E** and **H** vectors". Equations (44)/(42) cannot represent inductive mappings as they stand. The vectors $\mathbf{e}_j$ and $\mathbf{h}_j$ are normalized, so that both the left and right hand side of the equations have the same units, however, equation (44b) is impossible as an inductive process because $\bar{\mathbf{Z}}$ is not an admittance matrix. $\mathbf{Z}$ is not unitary, so that its adjoint will be its inverse, therefore, the equation cannot represent the inductive mapping of an electric vector onto a magnetic vector, even if they are normalized. $\mathbf{Z}$ is not even Hermitian so that the normalized $\bar{\mathbf{Z}}$ will become its inverse. In view of

the rotation group definitions, the answer as to what they represent, is quite simply coordinate mappings; equation 44(b) in particular represents coordinate transformations on the image $\bar{Z}$.

Secondly, I find the 'miraculous' transition from the real singular value matrix (44c) to the complex form (42) unconvincing; I fail to see on what grounds the singular values become automatically complex by imposing the constraints (45a). Moreover, the parameter counting system used by the authors is incorrect. The authors construct unitary matrices that are supposed to facilitate inductive mappings. For any kind of mappings to be possible, the transformation operators must conserve some kind of norm, i.e. there must exist a measure of invariance. The first and most obvious constraint to consider is the determinant of the matrix operator. The authors failed to do this; in hindsight we know that they would have found that det[**U**] = 1, which imposes a third real constraint on **U**. Therefore, their system of parameter count produces eight 'constraints' in (44c) and ten constraints in (42). This system of parameter counting is clearly problematic. If it were correct, any manipulation of a system of matrix equations would disturb the balance of parameters, which is intuitively unwarranted. The authors do not define at all what they mean by the term 'parameter', or rather they assign to it the rather downrated significance of a real number. I believe that this is a result of their attempt to adapt the real SVD to the complex case, which is a theoretically non-trivial matter, and may lead to unwarranted assertions.

In the real SVD, when real rotation operators with n(n-1)/2 real parameters are involved, the counting system of LaTorraca et al is valid

and the equation $n^2 = n(n-1)/2 + n + n(n-1)/2$ gives the balance of constraints in the LHS and RHS of the decomposition. In the complex case however, things are much more complicated. Whereas for real nxn rotation operators one has to start with n^2 real parameters and use the orthogonality relationships to reduce them to the n(n-1)/2 *independent* constraints sufficient to define the matrix, (e.g. Wigner, 1959), in the complex case one has to start with $2n^2$ parameters, in which case the orthogonality constraints are not enough. This reflects the profound difference in the dimensionality of the space that can be described by real and complex operators. The parameter counting system must be understood within the the context of the vector space defined by the rotation operator. Thus, in the case of 2x2 real rotation operators, this is the 2-D real basis E_2, and the one degree of freedom says that a rotation can only follow one path. There exists only one 2x2 real rotation operator with respect to which the decomposition is unique. In the case of 2x2 unitary rotation operators however, the vector space is 3-D and this imposes the two additional constraints. The three degrees of freedom describe the possibility of rotations when three parameters (angles) vary simultaneously. As a consequence, there exists a whole group of candidate rotation operators, but only one decomposition, depending on the particular physics of the matrix we seek to decompose. Therefore, it is unwarranted to try to adapt the real SVD to a complex case, by using the same parameter counting system, because of the danger of overconstraining the operators. The complex equivalent of the real SVD either exists or not; if it exists, it must be defined within the context of the space involved; within this context, the proofs of Propositions 1 and 2 only show that such a decomposition is a possibility. A fine example can be afforded for the 3x3 case, which for

real matrices defines three degrees of freedom and a 3-D vector space, while for unitary 3x3 rotation matrices defines *eight* degrees of freedom and an 8-D vector space. The existence of such a space depends explicitly on the Lie algebra (the structure constants) of the the group S(8) generating the symmetries of SU(3). The structure constants of SU(3) obey commutation relationships similar to those of SU(2), i.e. only two members of S(8) are allowed to commute at each time, and leads to the 'eightfold way' (eightfold symmetries of Quantum Chromodynamics), of Gell'mann (Gell'mann and Nee'man, 1965). Therefore, 3-D subspaces can readily be defined, but one realizes that from the endless possibilities generated, only a few can fit the particular physics of a 3x3 regular complex matrix. Therefore, the choice of the operator must take into account the task it is required to accomplish. The parameter count will always even out itself, as shown in V.2.2. I would like to point out here that (44c) is also problematic, in view of the real 'singular value' matrix it contains. Assuming its validity however, the parameter count *is even* because the LHS is a product, and only its linearly independent factors can be counted separately. Each element of **Z** will be always given as a linearly independent combination of eight real numbers,in the same way detailed in V.2.2; this means that eight degrees of freedom will always be associated with the transformation. The same holds for equation (42), or indeed, any system of this kind. I believe that because of these problems, the theory of the authors' generalization of the SVD for complex matrices must be, partially at least, reworked, especially when square unitary matrices are involved.

Conditions (45a) are true for certain classes of rotation operators. They do not reduce the number of degrees of freedom in **U**, but they have a profound effect! The most obvious form of elliptically polarized

rotating vectors (e_j , h_j) obeying (45a) have the form (32), and the eigenvectors $V^{\pm}$ can indeed be put into the form (45b); this is another way of expressing (30) and (31) in the two-dimensional basis (x, y). Condition (45c) also follows from the interpretation of e_j , h_j (or equivalently $V^{\pm}$) as field vectors. Nonetheless, geometrically it is valid because the eigenvectors $V^{\pm}$ are the 'carrier' coordinates of the characteristic states. As will be seen in Section VI, the bounds (45d) are overestimated, because the half-angle parameterization of SU(2) is not taken into account.

In V.6.1, I indicated that the SVD, as a concept, is a physically implausible decomposition for Z. Moreover, I conclude that the formulation of LaTorraca et al, is empirical, and produces a correct conclusion (i.e. 45a), from a largely unwarranted assertion. Thus, I suggest that the identification of the authors' U' with the operators $U_z \bar{U}_y$ is largely coincidental. The question of why the results of the authors are valid can now be answered. This is because the SVD formulation will lead to expressions similar to (36a, b) for the C matrices, with non-increasing diagonal entries for both D'_E and D'_H . Once unitary matrices of the form (26) are used, the estimation procedure will result in small differences in the estimation of the β angles only. Also, conditions (45c,d) can only be fulfilled if the mappings $h_1(\beta'_H, \gamma'_H) \rightarrow e_1(-\beta'_E, \gamma'_E+\pi/2)$ and $h_2(-\beta_H, \gamma'_H+ \pi/2) \rightarrow e_1(\beta'_E, \gamma'_E)$ are made; these are also the results of equations (39). Therefore, the product $M' = \bar{U}_E \, Z \, [U_H \, s_1]$ will return a diagonal 'singular' value matrix. The Pauli reflection matrix s_1 is used to denote the interchanging of the (h_1, h_2) column vectors. Thus we can see that the SVD and the UD are not *independent*. They are related by reflection symmetry, and are, thus, defined in anti-symmetric (dual) vector spaces. In this respect

we can understand why the SVD provides valid estimates of the rotation angles, and, under certain estimation conditions that may result from an incorrect substitution of values in (26) (i.e. $\gamma'_H + \pi/2$ for γ'_H), or independent estimation of the 'singular values' the formulation of LaTorraca et al will produce the numerically correct result!

The reason I used an operator of the type (26) instead of the straight product $\mathbf{U}_{zy} = \mathbf{U}_z\mathbf{U}_y$, and the same notation as with the authors, is so that direct comparisons can be made. If one bothers to undertake the exercise, one will find that $\mathbf{U}_{zy}$ also has the form $\mathbf{U}_{zy} = [\mathbf{V}^+ \; \mathbf{V}^-]$. Because $\bar{\mathbf{U}}_y = \mathbf{U}_y(-\beta)$, the two operators $\mathbf{U}_z\mathbf{U}_y$ and $\mathbf{U}_z\bar{\mathbf{U}}_y$ are revealed to produce characteristic states that are *enantiomorphic* with respect to the (x,y) plane. The physical processes and the resulting parameter sets behave identically, and admit the same interpretations. More details will be discussed at a later stage.

VI. Parameter estimation

The estimation of the β and γ angles parameters, as well as the eigenvalues of **Z**, can be carried out in a manner similar to that of LaTorraca et al (1986). From the equations (36) we can see that

$$\text{Det}[\mathbf{C}_E] = \text{Det}[\mathbf{C}_H] = r_1^2\, r_2^2,$$

and

$$\text{Trace}[\mathbf{C}_E] = \text{Trace}[\mathbf{C}_H] = r_1^2 + r_2^2$$

which provide a system of equations that facilitate the computation of the eigenvalues of the **C** matrices. Make use of the form (26) and substitute in each of the equations (36a) and (36b), to obtain after some algebra the following

$$\mathbf{C}_E = \tfrac{1}{2}(r_1^2+r_2^2)\mathbf{I} + \tfrac{1}{2}(r_1^2-r_2^2)\cos 2\beta_E \times \begin{bmatrix} \cos 2\gamma_E & \sin 2\gamma_E - i\tan 2\beta_E \\ \sin 2\gamma_E + i\tan 2\beta_E & -\cos 2\gamma_E \end{bmatrix} \quad (46a)$$

and

$$\mathbf{C}_H = \tfrac{1}{2}(r_1^2+r_2^2)\mathbf{I} + \tfrac{1}{2}(r_2^2-r_1^2)\cos 2\beta_H \times \begin{bmatrix} \cos 2\gamma_H & \sin 2\gamma_H - i\tan 2\beta_H \\ \sin 2\gamma_H + i\tan 2\beta_H & -\cos 2\gamma_H \end{bmatrix} \quad (46b)$$

The expressions giving the four angles can be obtained by identifying the right hand side and the left hand side elements of $\mathbf{C}_E$ and $\mathbf{C}_H$ in (46); these are

$$\tan 2\gamma_E = \frac{C_E(1,2) + C_E(2,1)}{C_E(1,1) - C_E(2,2)} \qquad \tan 2\gamma_H = \frac{C_H(1,2) + C_H(2,1)}{C_H(1,1) - C_H(2,2)} \quad (47a)$$

and

$$\sin 2\beta_E = \frac{i(C_E(1,2) - C_E(2,1))}{(r_1^2 - r_2^2)} \qquad \sin 2\beta_H = \frac{i(C_H(1,2) - C_H(2,1))}{(r_2^2 - r_1^2)} \quad (47b)$$

Then, **M** can be computed directly by substitution of parameters in (26) and (36c). This procedure differs from that LaTorraca et al, only in the estimation of β_H. The difference is imposed by the requirement that **M** be antidiagonal.

The 'skew angle'

$$a = \tfrac{1}{2}\pi - \gamma_H + \gamma_E$$

first introduced by Eggers and discussed in more detail by LaTorraca et al, is an additional effective measure of three-dimensionality through its departure from $\pi/2$. Details can be found in the reference papers. Here, I recommend the simple difference $a' = \gamma_E - \gamma_H$ that will be an equivalent measure, through its departure from zero. I introduce the simplified version, bearing in mind that according to the analysis, the two intrinsic coordinate systems coincide in the 2-D approximation. The modified skew angle a' is bounded as $[-\pi/2, \pi/2]$ and its sign provides information about the relative orientation of the intrinsic coordinate systems.

The bounds imposed on β and γ are given by

$$- \pi/4 \leq \beta_E, \gamma_E, \beta_H, \gamma_H \leq \pi/4 \qquad (48)$$

This is a direct result of the half-angle definition of SU(2) and the rotational (twofold) symmetry of Z. Recall that the impedance polar diagrams on the (x,y) plane consist of four asymmetric leaves that repeat through rotations by radians. Thus, while the rotation space is $[-\pi/2, \pi/2]$ the SU(2) matrices accomplish the rotation by only spanning the space $[-\pi/4, \pi/4]$. In a physical coordinate system, the application of the UD rotations are implied to be due north on the (x,y) plane. Therefore, the bounds (48) on and must be taken in the sense $W-(-\pi/4) \leq \gamma_E, \gamma_H \leq (\pi/4)-E$. Thus, a note of caution applies, in connection with the actual azimuths of the intrinsic coordinate system and conditions (48). Let me again underline the fact that the half-angle parametrization of SU(2) by no means implies that the computed parameters β and γ have to be doubled. This is only a mathematical condition. The bounds (48) however, are real. Thus, if for example there exists such an intrinsic coordinate system that $|\gamma_E| > \pi/4$ and $|\gamma_H| > \pi/4$ the structure of the rotation operator **U** changes and orthogonality is maintained by γ_E and γ_H swapping quadrants (reversing sign). This is equivalent with the well known phenomenon of major and minor axis swapping in the conventional analysis procedures. Then, application of (36c) will return

$$\hat{\mathbf{M}} = \begin{bmatrix} a & \hat{\mu}_2 \\ \hat{\mu}_1 & b \end{bmatrix}$$

where a,b may be significant numbers, and $\hat{\mu}_j$ may not be extremal. Worse results must be expected if it so happens that *only one* of the physical coordinate system angles exceeds the limit of $\pi/4$. Then, because γ_E and γ_H are independently calculated, only one of the angles

swaps quadrants, and application of (36) produces a totally distorted and unreliable **M**. The chances of the β angles swapping quadrants are minimal because it may be argued that no real Earth can provide such a conductivity distribution that the characteristic states tilt so much. This would require EM field components with a vertical magnitude larger than that of the horizontal.

The effects of such situations provide the means for their correction. Firstly note that under the conditions of quadrant and sign inversion, the correct azimuths are provided by a symmetry of $\pi/2$. Secondly, let it be understood that if the estimation procedure is correctly applied, reduction of the diagonal elements of the 3-D impedance tensor to magnitudes of the order of, say, 10^{-3} is intolerable, and it indicates an inadequate decomposition. For single precision arithmetic in the computer system I use, a satisfactory result implies that the modulus of the diagonal elements $|M_{xx}|$ and $|M_{yy}|$ must be reduced to the order of 10^{-8} to 10^{-9}, depending on the quality of estimation of the tensor **Z**. The following steps will remedy all the above problems, and it is recommended that they be taken as a standard.

(a) Test if $|M_{xy}|=\max(|M_{xy}|, |M_{xx}|)$ and $|M_{yx}|=\max(|M_{yx}|, |M_{yy}|)$.

(b) Test if $|M_{xx}| > a$ and $|M_{yy}| > a$ with, say, $a = 10^{-3}$.

If both statements are false, no problem exists. If one of the statements is true proceed to the following corrections: Generate new matrices $\mathbf{U}'_E$ and $\mathbf{U}'_H$ by substituting for the angles $\gamma'_E=\gamma_E+\pi/2$, $\gamma'_H=\gamma_H+\pi/2$ in (36b). By adding $\pi/2$ to the estimated γ angles, the corrections always follow the movement of the physical azimuth. Note that in the 'corrected' coordinate system, the bounds (48) also apply. Next apply (36c) as

$$\mathbf{M}_1 = \bar{\mathbf{U}}'_E \, \mathbf{Z} \, \mathbf{U}_H$$

$$\mathbf{M}_2 = \bar{\mathbf{U}}_E \, \mathbf{Z} \, \mathbf{U}'_H$$

$$\mathbf{M}_3 = \bar{\mathbf{U}}'_E \, \mathbf{Z} \, \mathbf{U}'_H$$

Repeat the tests (a) and (b) for each of the $\mathbf{M}_1$, $\mathbf{M}_2$ and $\mathbf{M}_3$. The the following statements apply:

(1) If $\mathbf{M}_1$ satisfies the tests, $\mathbf{M}_2$ and $\mathbf{M}_3$ do not. The distortion was due to γ_E swapping quadrants and $\mathbf{M}_1$, is the correct estimate, associated with $\gamma_E + \pi/2$ and.

(2) If $\mathbf{M}_2$ satisfies the tests $\mathbf{M}_1$ and $\mathbf{M}_3$ do not. The distortion was due to γ_H swapping quadrants. Adopt the corrections.

(3) If $\mathbf{M}_3$ satisfies the tests $\mathbf{M}_1$ and $\mathbf{M}_2$ do not. The distortion was due due to *both* γ_E and γ_H swapping quadrants. Adopt the corrections.

If the matrix $\mathbf{U}_{zy}$ had been used instead of the form (26), the difference in (46) would have been a simple *transposition*. Thus from (47a) we see that the γ angles would be estimated identically, while from (47b) the β angles would change sign, implying the existence of enantiomorphic states, with respect to the (x,y) plane. In terms of 3-D rotations, one must note that there is a penalty to be paid for attempting the rotation without measurements in the vertical direction. Thus, we can only estimate the magnitude of the required rotation, but we are unable to constrain the direction of inclination (β) of the resulting states, without additional (e.g. tipper) information. From a mathematical point of view alone, i is impossible to decide which rotation is consistent with the geoelectric structure producing the measured impedances, and as a consequece, the handedness of the rotation of the the EM fields; only their magnitudes are unambiguous. In order to obtain handedness information, measurements in three-dimensions would be needed, so that the determination of β would be unambiguous.

VII. Comparisons with other analyses.

Comparisons are now presented of the performance of the UD and the other methods found in the reference papers quoted. All comparisons are based on parameter estimation for the example impedance tensor given by Eggers (1982)

$$\mathbf{Z} = \begin{bmatrix} 0.097 + i0.208 & 1.140 + i0.957 \\ -0.274 - i0.457 & 0.297 - i0.138 \end{bmatrix}$$

which is already rotated to the conventional principal directions. The parameters estimated by the UD and shown in Figure 2a, bear the same notation as in the text. The parameters associated with the maximum state (E^+, H^-, $\mu^+ \equiv \mu_1$) are shown on the left, while the corresponding minimum state (H^+, E^-, $\mu^- \equiv \mu_2$) parameters are shown on the right. Because a three-dimensional representation of the the UD results would be rather complicated, I opted to produce the characteristic states on the (**x**, **y**) basis, where the angles β are displayed as ellipticities. The explicit result of the product (36c) is given by the matrix

$$\mathbf{M}_{UD} = \begin{bmatrix} -1.490\text{x}10^{-8} - i7.450\text{x}10^{-9} & 1.178 + i0.988 \\ -0.310 - i0.435 & 2.980\text{x}10^{-8} - i1.490\text{x}10^{-8} \end{bmatrix}$$

With the main theoretical details discussed, no results of the analysis of LaTorraca et al (1986) are presented, because their numerical values are identical to the UD.

VII.1 Comparisons with Eggers' (1982) biorthogonal analysis.

In order to extract anti-diagonal principal components for the impedance tensor, Eggers (1982) based his eigenstate formulation on the assumption of TEM waves propagating into the Earth. Also, Spitz (1985) worked on

an implicitly similar basis. Specifically, Eggers defined the impedance tensor eigenstates through the expressions

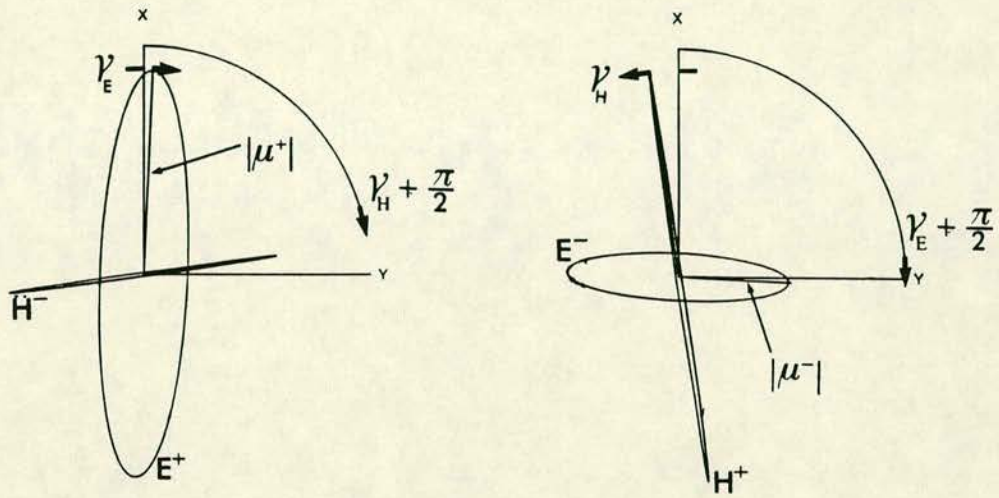
$$\mathbf{E} = \underline{\Lambda}^i \mathbf{H} \tag{49a}$$

$$\mathbf{E}\,\mathbf{H} = 0 \tag{49b}$$

(49b) is the orthogonality property between fields in the TEM mode. This formulation requires the magnetic field state vector **H** and its linear transform **E** to be orthogonal in the generalized vector space; the necessary and sufficient condition for (49) to hold is that $\underline{\Lambda}^i$ be anti-symmetric. The pairs of eigenvectors ($\mathbf{H}^i$, $\mathbf{E}^i = \mathbf{Z}\mathbf{H}^i$) are at right angles because

$$\mathbf{E}^i = \underline{\Lambda}^i \mathbf{H}^i = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \lambda^i \mathbf{H}^i = \underline{\theta}(\pi/2)\, \lambda^i \mathbf{H}^i$$

where $\underline{\theta}(\pi/2)$ denotes the $\pi/2$ rotation operator. As Eggers indicates, these conditions do not require the two eigenstates to be perpendicular in space, except for the 2-D approximation. Furthermore, the eigenvectors $\mathbf{E}^i$ and $\mathbf{H}^i$, defined in the basis (29) with expressions similar to (30) represent elliptically polarized fields. However, the conditions (49b) are somewhat artificial, and biorthogonal analysis is still confined to the two-dimensional space of the horizontal plane; it does not consider the three-space at all. This has unfortunate effects on the estimation procedure. A more meticulous analysis of the eigenstate formulation by Spitz (1985) reveals that if **Z** does not become traceless, through a rotation by π radians, or if there does not exist an angle α_0 such that one of the diagonal elements vanishes for $\mathbf{Z}'(\alpha_0)$, the standard eigenvalue expressions do not stand for the off-diagonal elements of a **Z** measured in a physical coordinate system. Therefore, the eigenstate formulation is not always guaranteed to return the extremal values of the impedance. Figure 2b, illustrates the

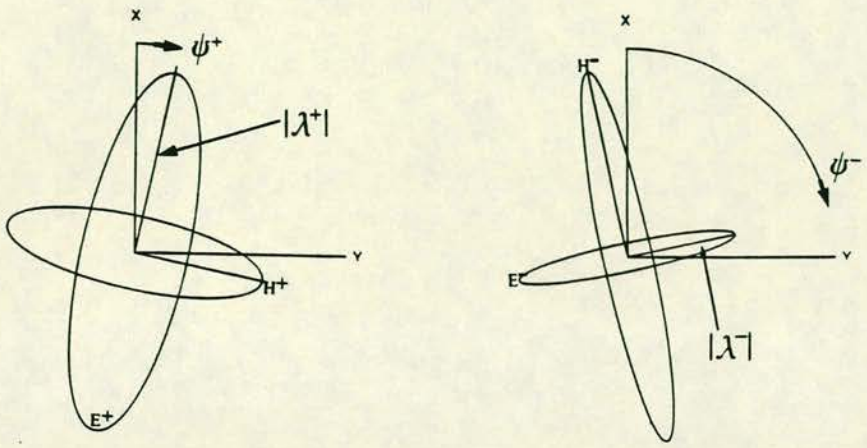


$\mu^+ = 1.538\exp(i39.9°)$ $\quad$ $\mu^- = 0.542\exp(i54.1°)$

$\gamma_E = 2.58°$ $\quad$ $\gamma_H = -8.18°$

$\beta_E = -12.24° \left(e_E = -0.217\right)$ $\quad$ $\beta_H = 0.81° \left(e_H = 0.0141\right)$

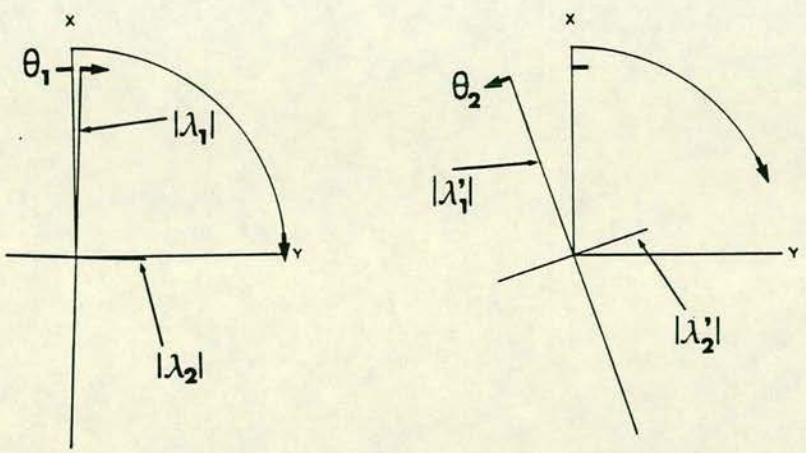
FIGURE 2 (a) The UD analysis of Eggers' (1982) example impedance. The two characteristic states are produced by a unit amplitude linearly polarized magnetic field at the orthogonal directions (β_H, γ_H) and $(\beta_H, \gamma_H + \pi/2)$. The dip angles β_E and β_H are shown in the form of ellipticities only.



$\lambda^+ = 1.420\exp(i41.7°)$
$\psi^+ = 12°$
$e^+ = -0.288$

$\lambda^- = 0.588\exp(i52.9°)$
$\psi^- = 72°$
$e^- = -0.130$

FIGURE 2 (b) Eggers' biorthogonal analysis for his example impedance tensor. The two states are generated by unit amplitude magnetic fields.



$\lambda_1 = 1.48\exp(i39°)$
$\lambda_2 = 0.543\exp(i61°)$
$\theta_1 = 2.55°$

$\lambda_1' = 1.41\exp(i46°)$
$\lambda_2' = 0.586\exp(i43°)$
$\theta_2 = -18.8°$

FIGURE 2 (d) The results of Spitz's analysis on Eggers' example impedance tensor. Spitz's approach does not provide a formal definition of ellipticity, therefore the 'states' are presented in an amplitude form only.

performance of Eggers' biorthogonal analysis. For the example impedance, there exists no real angle such that one of the diagonal elements vanishes on the (x,y) plane, therefore, the eigenstate formulation is insufficient to provide the extremal values of the impedance. It is clear that the eigenvalues computed through Eggers' eigenstate formulation lie somewhere in between the extremal values obtained with the UD. Substitution of the angles computed by the bi-orthogonal analysis (horizontal azimuths ψ^+ and ψ^-, and the arctangents of the ellipticities e^+ and e^-) in (26b) and (36c) gives

$$\mathbf{M}_{\text{biorthogonal}} = \begin{bmatrix} 2.605\text{x}10^{-1} - i7.450\text{x}10^{-2} & 1.201 + i0.877 \\ -0.269 - i0.414 & -0.312\text{x}10^{-2} - i6.031\text{x}10^{-1} \end{bmatrix}$$

Naturally, $\mathbf{M}_{\text{biorthogonal}}$ does not contain the same values, returned by the eigenvalue expressions used by Eggers (Figure 2b), that are the correct results of the analysis. However, because the parameters are theoretically equivalent, this exercise helps to demonstrate that the values returned by biorthogonal analysis are not sufficient to anti-diagonalize the example impedance tensor.

VII.2 On the question of an orthogonal decomposition for Z.

It would be reasonable to ask whether simple orthogonal transformations of the form (11) would be sufficient to diagonalize **C**. The answer is negative. The matrices **C** can of course be rotated by orthogonal operators, but they are subject to similar topological constraints, as the rotation of the the impedance tensor **Z**. **C** comprise four degrees of freedom, while the rotation depends on three degrees of freedom. C_i, i=1,..,4 all comprise 1 degree of freedom according to the definition (12), and the azimuth obtained by maximizing C_3 depends on two degrees

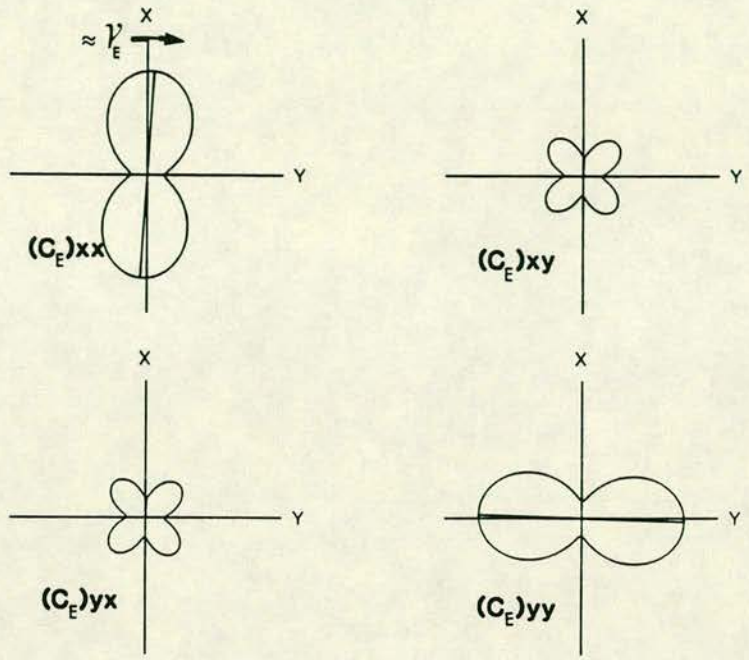
of freedom only. Furthermore, the rotation of the diagonal and anti-diagonal elements is still independent by C_1 and C_2. An incomplete set of parameters results, and the rotation is not guaranteed to diagonalize **C**. In general, it will find maximum values that, due to the reduced dimensionality of the **C** parameter space, will be much more succesfull in locating the extremal values than the corresponding rotation of **Z**. Figure 2c shows the polar diagrams for the rotations of $\mathbf{C}_E$ and $\mathbf{C}_H$. It can readily be seen that $\mathbf{C}_H$ virtually diagonalizes at the principal directions (-8.72°, 82.18°). However, $\mathbf{C}_E$ does not diagonalize at the principal directions (2.58°, 92.58°), e.g. $(\mathbf{C}_E)_{xx}$ and $(\mathbf{C}_E)_{yy}$, although the angles are estimated within a fraction of the extremal values $(\gamma_E, \gamma_E+\pi/2)$ located by the UD. From Figure 2a, one can see that β_E is substantial (12.3°), while β_H is very small (0.81°). This suffices to explain the differences in the rotation polar diagrams. From a mathematical point of view, a complex matrix with four-dimensional topology is needed to perform the task. From a physical point of view, **C** contains the symmetries of **Z**, and the angle β is is needed as well in order to unfold them. It follows that no decomposition of the form

$$\mathbf{Z}'(\gamma_E, \gamma_H) = \mathbf{R}^T(\gamma_E)\ \mathbf{Z}\ \mathbf{R}(\gamma_H) \qquad (50)$$

can be furnished, such that **Z'** is antidiagonal, because the coupling between the vertical components is not removed. Nonetheless, if the 3-D influences are small, such a decomposition will 'nearly miss' the extremal values of **Z**. Application of (50) with the γ angles will return

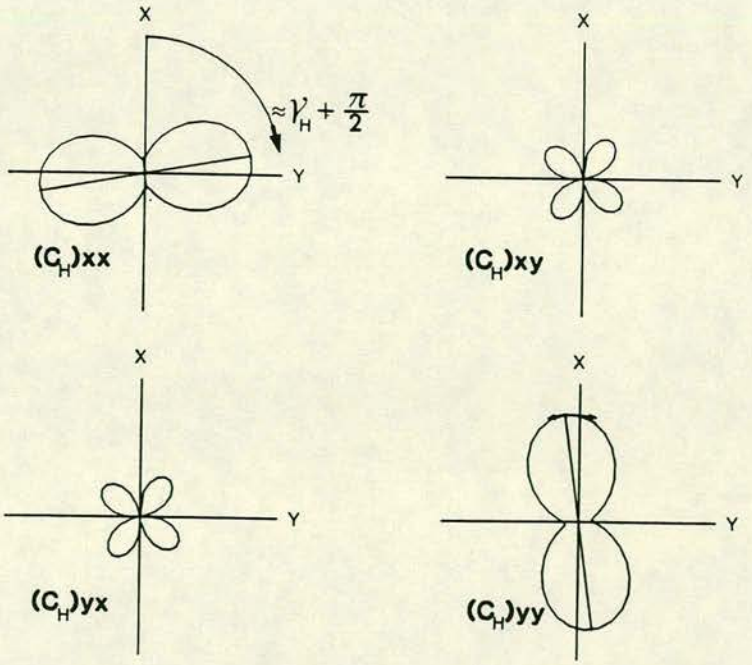
$$\mathbf{Z}' = \begin{bmatrix} -8.015\text{x}10^{-2} + i5.023\text{x}10^{-2} & 1.153 + i0.967 \\ -0.310 - i0.435 & 0.203 - i0.245 \end{bmatrix}$$

with $Z'_{xy} = 1.504\ \exp(i39.98°)$, and $-Z'_{yx} = 0.534\ \exp(i54.5°)$. This decomposition provides the second best performance in terms of locating



(1)

FIGURE 2 (c) Polar diagrams of the conventionally rotated matrices (1) $C_E = Z\,\overline{Z}$, and (2) $C_H = \overline{Z}\,Z$, for Eggers' example impedance tensor.



(2)

the extremal values of the example impedance. Nonetheless, the tensor does not reduce to a zero-diagonal form. The stronger the 3-D effects become, the more decomposition (50) will deviate from a zero-diagonal form. It is worthwhile to notice that a decomposition of the form (50) shares the philisophical basis, while avoiding the elaborate mathematics of the work of Counil et al (1986), who have provided a complete theory of tensor tranformations that map the horizontal electric field onto the horizontal magnetic field and vice versa. In their work, the orthogonal 'extremal' directions γ_E and γ_H will be found under the terminology of 'electric and magnetic associate and conjugate directions' respectively.

VII.3 Comparisons with Spitz's Cayley's factorization approach.

Spitz's (1985) paper presents a thorough and outstanding, albeit tedious exploration of the behaviour of the impedance tensor under conventional rotations. The author proposes two solutions to the problem; I shall not attempt to describe his analysis and the first solution, since it has several non-trivial features that hinder a brief outline. This was intended as a demonstration of properties and not as an estimation procedure, inasmuch it possesses non-unique solutions. It is however a commendable and valuable source of information. I will concentrate on the second solution the author proposes, that involves a Cayley factorization of the impedance tensor. This comprises the decomposition of a regular complex 2x2 matrix as

$$\mathbf{X} = \mathbf{Y}\,\mathbf{P} \qquad \text{and} \qquad \mathbf{X} = \mathbf{Q}\,\mathbf{W} \qquad (51)$$

where $\mathbf{Y}$ and $\mathbf{W}$ are unitary and $\mathbf{P}$ and $\mathbf{Q}$ are the positive definite solutions of the equations

$$\mathbf{P}^2 = \bar{\mathbf{X}}\,\mathbf{X} \qquad \text{and} \qquad \mathbf{Q}^2 = \mathbf{X}\,\bar{\mathbf{X}} \qquad (52)$$

This factorization is similar to the factorization of a complex number- the positive definite part plays the role of amplitude, and the unitary part plays the role of phase. Spitz uses the second of equations (51) and proposes to apply the conventional analysis technique in order to extract one angle θ_1 such that $|Q_1(\theta_1)|=\min$, that diagonalizes the positive definite matrix **Q**, and one angle θ_2 such that $|W_3(\theta_2)|=\min$, that antidiagonalizes the unitary part **W**. So, when **Z** is referred to the intrinsic coordinate system, or, when the conductivity distribution is 2-D,

$$\mathbf{Z}(\theta_1,\theta_2) = \begin{bmatrix} |\lambda_1| & 0 \\ 0 & |\lambda_2| \end{bmatrix} \begin{bmatrix} 0 & \exp(i\phi_1) \\ -\exp(i\phi_2) & 0 \end{bmatrix}$$

where $\lambda_j = |\lambda_j|\exp(i\phi_j)$ are the eigenvalues of **Z**. The notations Q_1 and W_3 imply the same quantities as in the expressions (12).

Now, consider that in order to extract a positive definite **Q** from (52) one has to diagonalize the Hermitian matrix $\mathbf{Z}\bar{\mathbf{Z}}$. This exercise has already been undertaken, using the unitary matrix $\mathbf{U}_E$. Thus, I write $\text{diag}[\mathbf{Q}] = \sqrt{\mathbf{D}_E}$. The conventional rotation of $\mathbf{Z}\bar{\mathbf{Z}}$ has been discussed above, and we know that the minimization of $|Q_1|$ locates the angle γ_E. The next step however, is problematic. The unitary 'phase' factor **W** has an eight-dimensional topology, since the phases of all Z_{ij} elements of the tensor should be expected to be different at an arbitrary orientation with respect to the principal directions. Its conventional rotation is subject to the same constraints as the rotation of **Z**. Thus, while the idea is correct, I do not see why the conventional analysis should be more efficient in anti-diagonalizing the unitary matrix **W**, than it is for the complex matrix **Z**. Spitz suggests that it is impossible to determine, from a mathematical point of view alone, which one of the θ_1 and θ_2 angles is best suited for MT data analysis. I

think it can be conclusively decided, that it is θ_1. It is probably a cruel irony that Spitz did not consider the alternative factorization. In this case, it is straightforward to verify that diag[P] = $\sqrt{D_H}$ and the application of Spitz's procedure to minimize $|P_1(\theta_1')|$ yields $\theta_1' = \gamma_H$. the angle θ_2' is again as problematic.

The angles θ_1 , θ_2 are shown as azimuths only, in Figure 2d. However, Spitz does not use Cayley's factorization to compute the characteristic values of **Z**. He bases his estimation procedure on the fact that *any* 2x2 complex matrix can be writen as a linear combination in terms of the Pauli matrices. Thus, **Z** can be writen as

$$\mathbf{Z} = Z_2\,\mathbf{I} + Z_3\,\mathbf{s}_1 + iZ_1\,\mathbf{s}_2 + Z_4\,\mathbf{s}_3$$

For a 2-D earth there exists one angle $\theta_0$such that **Z** reduces to the form (9), where the off-diagonal elements stand for its eigenvalues. In a 3-D Earth however, **Z** does not become zero-diagonal, and at any angle θ, it may be writen as

$$\mathbf{Z}(\theta) = Z_2\mathbf{I} + \begin{bmatrix} 0 & \lambda_1 \\ -\lambda_2 & 0 \end{bmatrix} + Z_4(\theta)\mathbf{s}_3 \qquad (53)$$

If $Z_4(\theta)$ is known, the off-diagonal elements can be put in the form

$$\lambda_{1,2} = Z_1 \pm Z_3(\theta) \qquad (54)$$

Spitz rotates the tensor conventionally to the directions θ_1/θ_2, and argues that in this way, it reduces to the form (53). Thus, he uses (54) to calculate the eigenvalues of the impedance.

For $\theta = \theta_1$, these give $\lambda_1 = 1.153 + i0.934$, and $\lambda_2 = -0.263 - i0.475$.

For $\theta = \theta_2$, they give $\lambda_1' = 0.981 + i1.016$, and $\lambda_2' = -0.430 - i0.401$.

Apparently, for the angle θ_1, the eigenvalues $\lambda_{1,2}$ are comparable to those of the other methods. Note however, that the phase of λ_2 is rather large, and, that the extremal values are not located either. The modulus of the eigenvalues does not correspond to the r_j obtained by

diagonalizing $\mathbf{Z}\,\bar{\mathbf{Z}}$. The parameters computed for the angle θ_2 are quite obviously in disagreement with any other method; I believe this demonstrates, empirically at least, that the angle θ_2 is not a principal direction.

VII.4 Recapitulation.

Eggers' (1982) paper, appears to have initiated a period of considerable effort to understand the properties of the impedance tensor in multidimensions. In addition to the investigations cited hitherto, other commendable papers are due to Counil et al (1986), and Zhang et al (1987). However, all the authors approach the subject from widely different perspectives, and none has actually presented a full theory of tensor transformations to describe the properties of **Z**. Counil et al (1986) present a proper tensor transformation theory, which however is still confined in the (x,y) plane, and therefore spatially incomplete. Zhang et al (1987) consider an impedance decomposition, that applies only to a particular model of a conductivity distribution, and, by the authors admission, can be quite problematic in different situations.

The wealth of information accumulated by all the above authors inspired the attempt to generalize the existing 2-D theory of impedance transformations. In my attempt to do so, I made use of the three-dimensional unitary rotation group SU(2), which comprises 2x2 matrices, and introduced a decomposition (the Unitary Decomposition), that transforms the experimental Cartesian frame of reference of the impedance tensor, to two Cartesian coordinate frames (the intrinsic coordinate system) that carry orthogonal and linearly polarized electric

and magnetic plane waves. The two frames are not necessarily orthogonal and horizontal; they become so in the limiting case of a 2-D geometry. The non-horizontal nature of the intrinsic coordinate system allows a quantitative estimation of the degree of distortion (coupling) experienced by the horizontal EM field components, due to the existence of the vertical components. This has been found to be related to the ellipticity of the EM field. Because unitary rotation operators are formally tensors, a complete theory of tensor transformations of **Z** in multidimensions, is provided.

The decomposition has been shown to yield the extremal values of **Z** under all circumstances, and to perform in a manner superior to the conventional analysis and to the other methods considered, providing a complete spatial description of **Z**. The performance of the method can also be demonstrated through its application to real MT sounding data, which is the subject of Part 3 of this chapter.

CHAPTER 1

AN ANALYSIS OF THE MAGNETOTELLURIC OPERATORS IN THREE DIMENSIONS

PART 3:

APPLICATION OF THE UD TO MAGNETOTELLURIC SOUNDING DATA

I now proceed to give some examples of the versatility of the UD by applying it to a set of real EM sounding data. It is true that a more conventional approach would be the application of the analysis to synthetic data, obtained by modelling a test 3-D conductivity distribution. Unfortunately, I do not have a 3-D modelling program. But even so, I still have doubts about how much reality we can generate with such a facility, given the computational effort it generally requires. Published models tend to be oversimplistic in their representation of a 3-D Earth. Usually, only simple regular shapes (e.g. rectangular prisms), embedded in otherwise isotropic matrices are used. The most complex cases found in literature, are the 'rectangular basin offsets' of Hermance (1982), and the 'L' shaped conductive valleys of Park et al (1983). Even these models have regular shapes, with the undesirable property that they display many planes of symmetry, near which the synthetic impedances simulate a nearly 2-D behaviour. I believe that although modelling offers excellent control, it is somewhat restrictive in interpretation, because it may be misleading, or, be approached with prejudices.

The data used in the present investigation have been obtained and described by Beamish and Banks (1983) and Banks and Beamish (1984), in N. England and S. Scotland; and are exhaustively investigated by Beamish (1986a,b). The paper of Beamish (1986b) focuses on the problem of inference of geoelectric structural dimensionality from MT data and

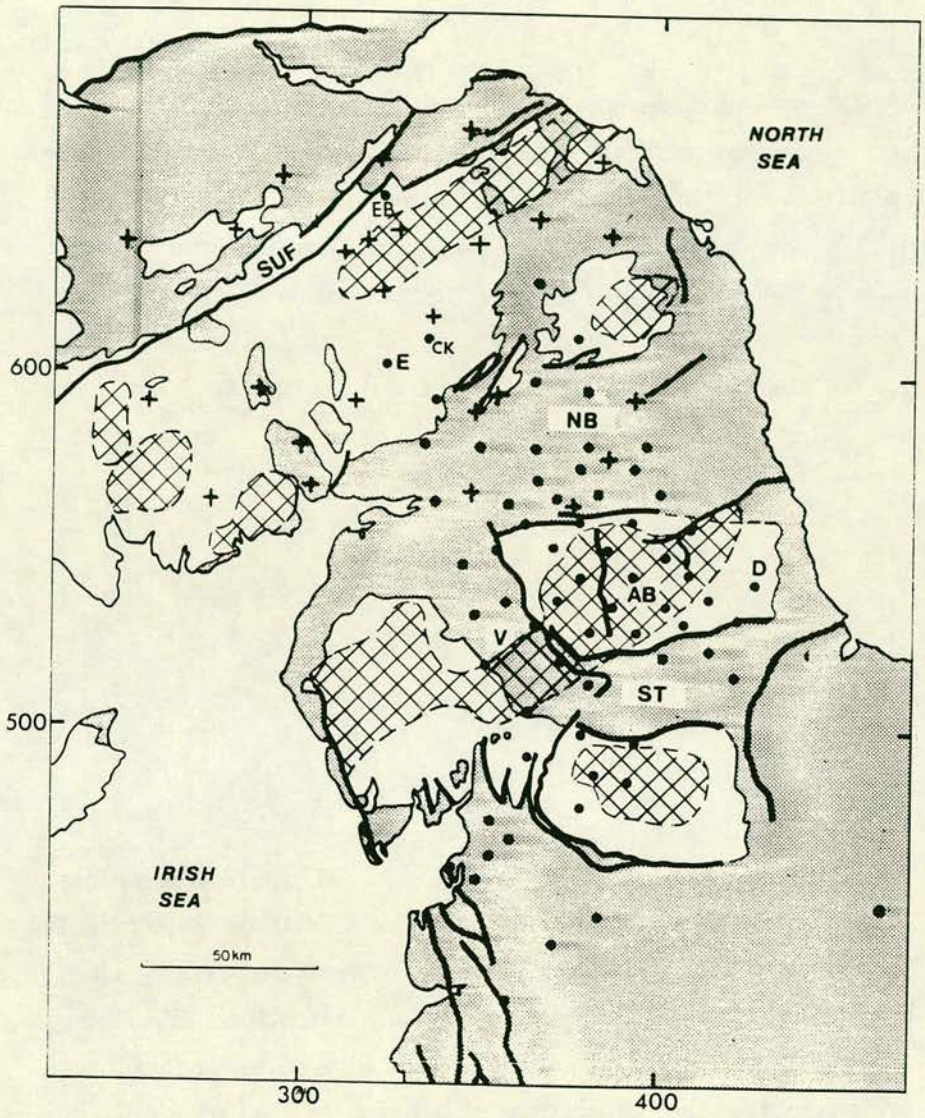
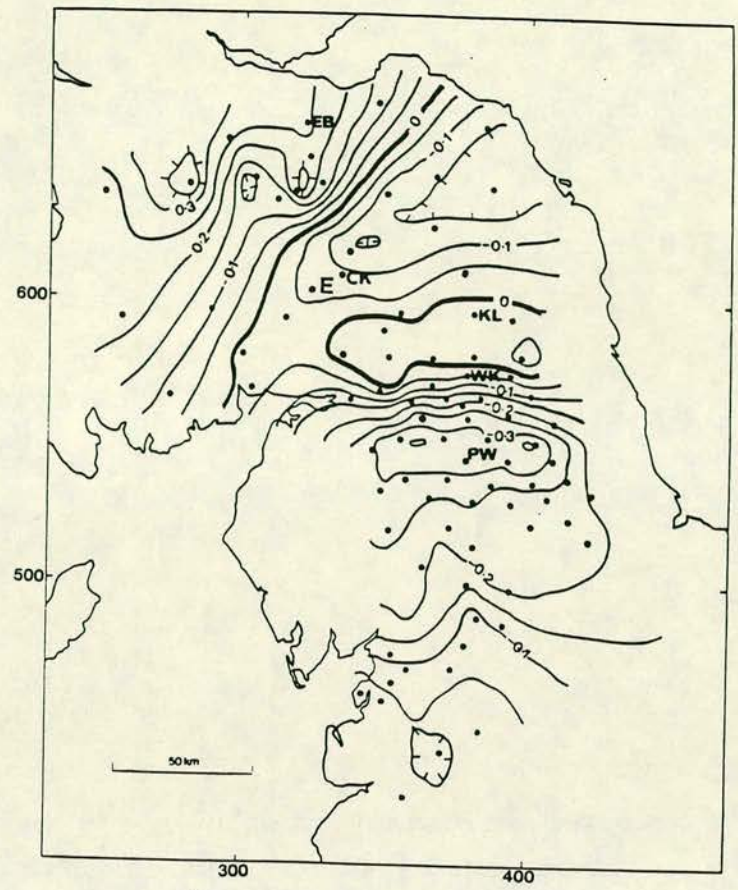


FIGURE 3 (Reproduced from Banks et al, 1983); Major geological features of N. England and S. Scotland. The NE-SW trending lineaments mark the closure of the Iapetus (proto-Atlantic) Ocean during the Caledonian Orogenesis. The thick lines are major faults. Hatched areas indicate regions underlain by igneous batholiths. Stippled areas are underlain by thick post-orogenic sediments. SUF, Southrern Uplands Fault; NB, Northumberland Basin, filled with relatively conducting Carboniferous sediment, it is suggested to mark the actual surface manifestation of the Iapetus suture; AB, Alston Block; V, Vale of Eden, filled with relatively conducting sediments of the Carboniferous-PermoTriassic (maximum thickness 3Km); ST, Stainmore Trough.

FIGURE 4 (Reproduced from Banks et al, 1983); map of the in-phase part of the vertical magnetic field produced when the unit amplitude horizontal field is directed towards magnetic north (9°W), at a period of 750sec. Expected direction of the current flow is E-W and should produce no vertical magnetic field if the stucture was homogeneous. Therefore, the vertical field observed is entirely due to horizontal conductivity contrasts.

FIGURE 5 (Reproduced from Banks et al, 1983); the anomalous current stream function in a thin sheet at a depth of 5Km, derived from Figure 5. Approximately 10A flow between the stream lines when a unit amplitude horizontal field is applied. The linear feature of the current in S. Scotland is similar to, but does not coincide with the Caledonian structures or any known belt of conductive rocks near the surface; instead, it appears to lie along the shortest conducting path that links the North and Irish Seas. The same can be observed for the Northumberland Basin. The Vale of Eden to the W of the Alston Block does not make such a link, insulated at its southern end by a granitic ridge, or by westward thining of sediments in the Stainmore Trough and does not respond to excitation by the unit amplitude field for the period 750sec. It provides a very probable source of 3-D effects.


EB
E
CK
KL
WK
PW
0
-0.1
-0.2
-0.3
600
500
50 km
300
400

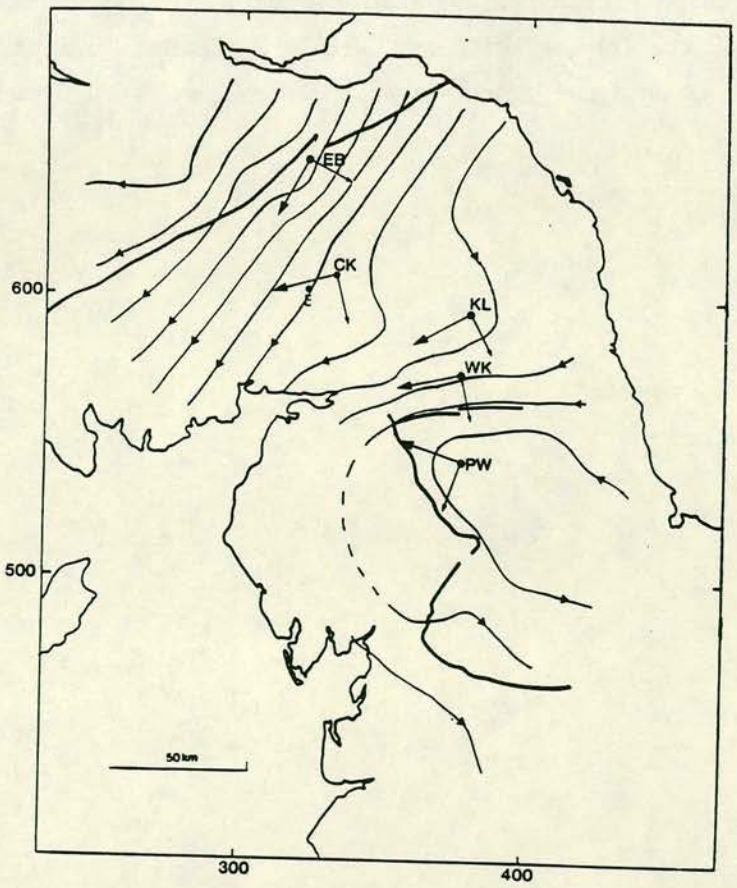
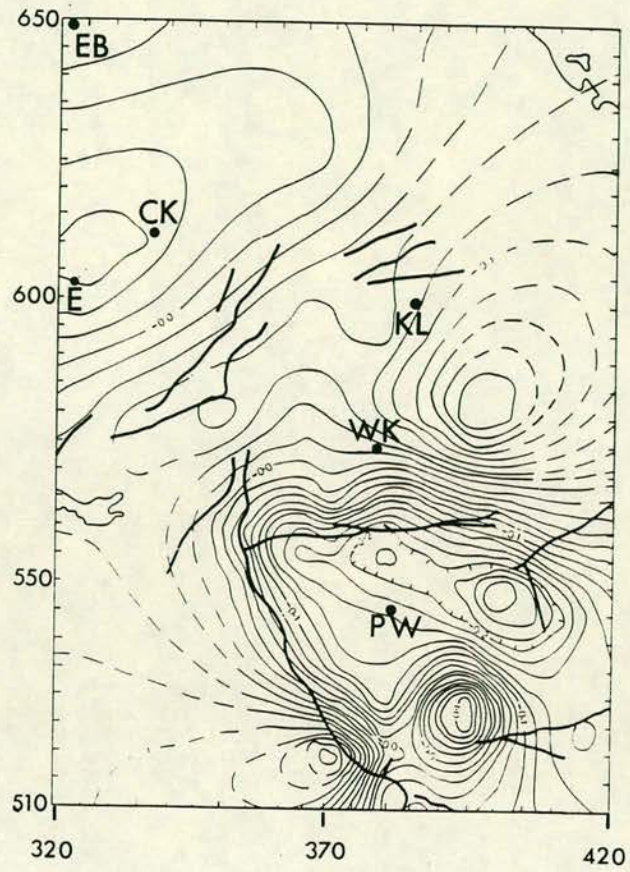
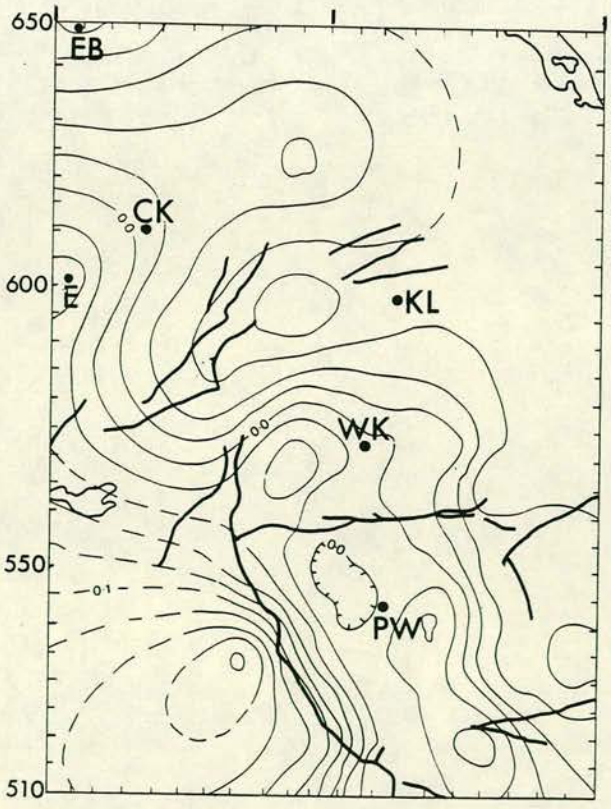

EB
CK
KL
WK
PW
600
500
50 km
300
400

FIGURE 6 (Reproduced from Banks and Beamish, 1984). Map of the in-phase part of the vertical (anomalous) magnetic field due to a unit amplitude horizontal component directed at **(a)** magnetic north (N9°W), and **(b)** magnetic east (N81°E). Contour interval is 0.02nT. Heavy lines indicate major fault systems. Period T = 60s.


650
600
550
510
320
370
420
EB
CK
E
KL
WK
PW
-00
-01

(a)


650
600
550
510
EB
CK
E
KL
WK
PW
00
01

(b)

thoroughly evaluates the established methods of estimation. I will make use of the data from sites 1 (EB), 2 (PW), B (KL) and C (WK), investigated by Beamish (1986b). The same sites with designations 1 (EB), 3 (KL), 4 (WK) and 5 (PW) appear in Beamish (1986a) in his exploration of the deep crustal structure beneath the Northumberland Basin (N. England). An outline of the large scale structural features of the N. England and S. Scotland area is given in Figure 3. More detailed geological information can be found in Legget et al (1983). Electrical models of the S. Scotland area can also be found in Ingham and Hutton (1982a,b). An overview of the anomalous fields associated with lateral structures across the whole region is given in Figures 4 and 5, reproduced from Banks et al (1983). The figures show the in-phase vertical magnetic field and associated current stream function for a period of 750sec, using the method of hypothetical event analysis from 94 GDS locations. Basic information is provided in the figure captions; detailed interpretations and comments can be found in Banks et al (1983). Figure 6 displays vertical anomalous magnetic fields due to a northward (Fig. 6a), and an eastward (Fig. 6b) unit amplitude horizontal magnetic field, at the relatively short period of 60sec. This information is reproduced from Banks and Beamish (1984). I now proceed to display and discuss the results of the UD for the aforementioned MT sites.

VIII. Sites WK and KL

Results from the site KL are presented in Figures 7a and 7b for the conventional analysis and the UD respectively; the skew angle is not shown. Figures 8a and 8b are the corresponding results for site WK.

More pictorial information is presented in the diagrams of Figures 9a (site KL) and 9b (site WK), that display the period (T) dependence of the characteristic states, in what I call the 'state mode' of presentation.

Beamish (1986a,b) found that these sites display 'normal' characteristics, at least for the bandwidth up to 400sec, in terms of all the conventional dimensionality indicators, and the existence and construction of 1-D solutions to the inverse problem. Thus, the presentation of results from these sites aims to demonstrate that the conventional analysis and the UD become equivalent when the dimensionality of the geoelectric structure is reduced. We observe that the azimuths γ_E and $\gamma_H+\pi/2$ are almost orthogonal for the periods $<$ 500sec, which indicates that the two characteristic states coincide. The largest observable deviation is of the order of 2.5°; deviations are unavoidable as a consequence of the uncertainties accumulated during data acquisition and processing.

If we first turn our attention to site KL (Fig. 7), we observe a very stable azimuthal direction for both γ_E and γ_H. However, for periods $>$ 1000sec the azimuth becomes less stable. KL is associated with linearly polarized fields, with the exception of the short periods (Fig. 9a). Two of the long period determinations are unstable and therefore of questionable reliability, despite the small statistical error they contain. One can readily observe that similar information is also provided by the conventional skew and ellipticity quantities. However, whereas the UD angles are physical and quantitatively interpretable, the skew and ellipticity only have a qualitative interpretation value. Moreover, in contrast to the conventional analysis, the UD provides

information about the magnetic field states. The significance of this will be discussed later. For this '1-D' site, the conventional and UD apparent resistivities display differences too small to be considered significant.

We next turn our attention to site WK (Fig. 8 and Fig. 9b). We observe that the azimuthal directions for both γ_E and γ_H are stable and orthogonal for periods < 500sec. For longer periods γ_H begins to rotate and β_H deviates significantly from zero. Orthogonality is marginally maintained, while neither γ_E rotates nor β_E moves away from 0°. For the period 750sec, the direction γ_E agrees well with the regional current system depicted in Figure 5. The conventional azimuth θ_c is almost identical to γ_E and for T < 1000sec behaves in a similar way; the shape of β_H and the conventional ellipticity curves are very similar, however the same comments as per site KL apply. The conventional rotation successfully locates the major and minor axes, which, as can be seen in Figure 8, are almost identical to the extremal values located by the UD for T < 1000sec.

If we study the results from both sites (Fig. 9), we observe that at short periods (T < 100sec) both sites show elliptically polarized states. From Figure 6a,b one observes that the structure around both sites responds to a horizontal magnetic field and produces an anomalous vertical component. The response is profoundly different at the two sites, and this leads one to conclude that the character of the observations in this period band, is due to local induction effects. For T > 100sec, KL certainly displays linearly polarized fields, while WK is associated with elliptically polarized H-fields; the electric components (arrowed) display a more stable linear polarization. With

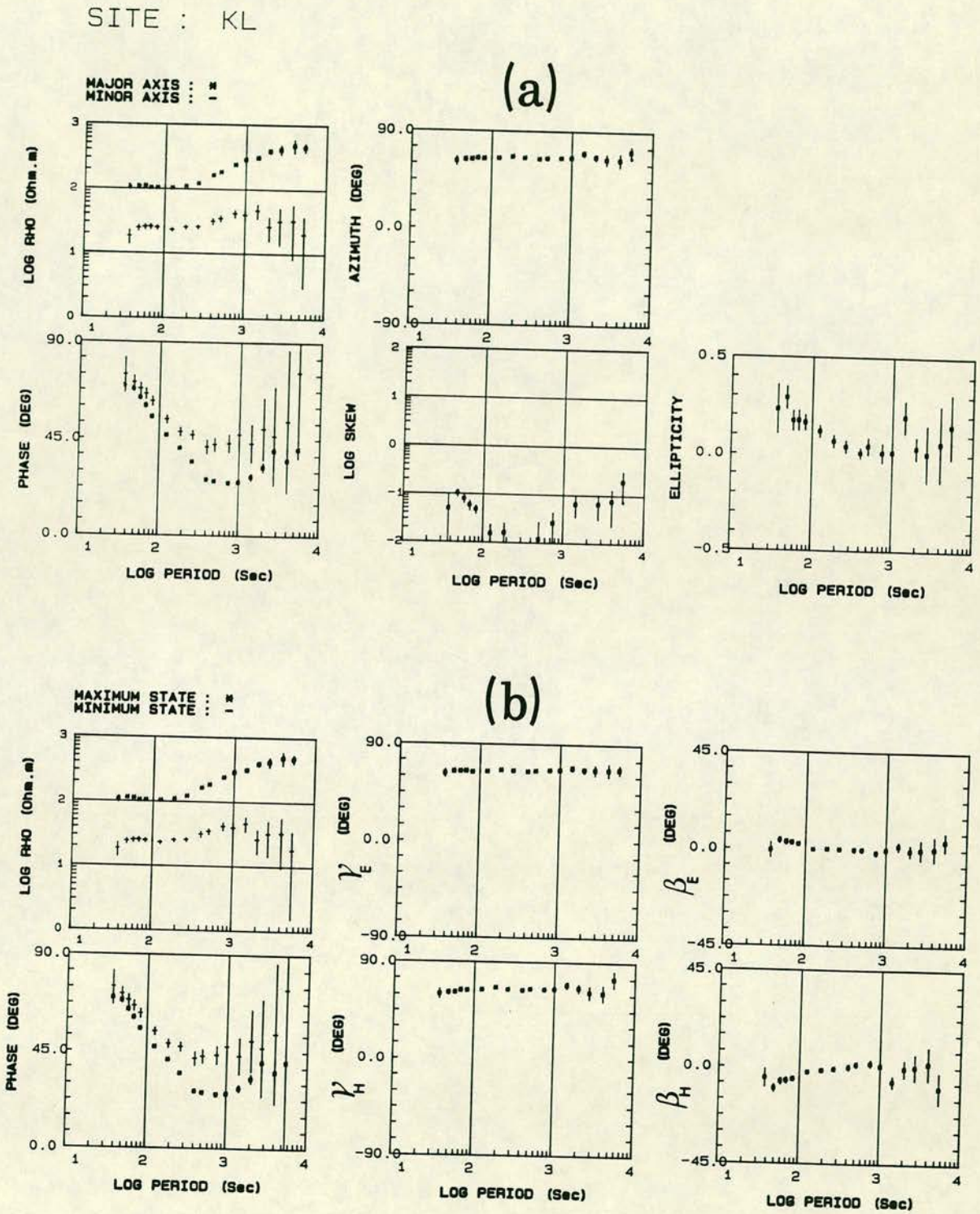


FIGURE 7 **(a)** The conventional analysis results for site KL.
(b) The UD results for site KL; the skew angle is not shown.

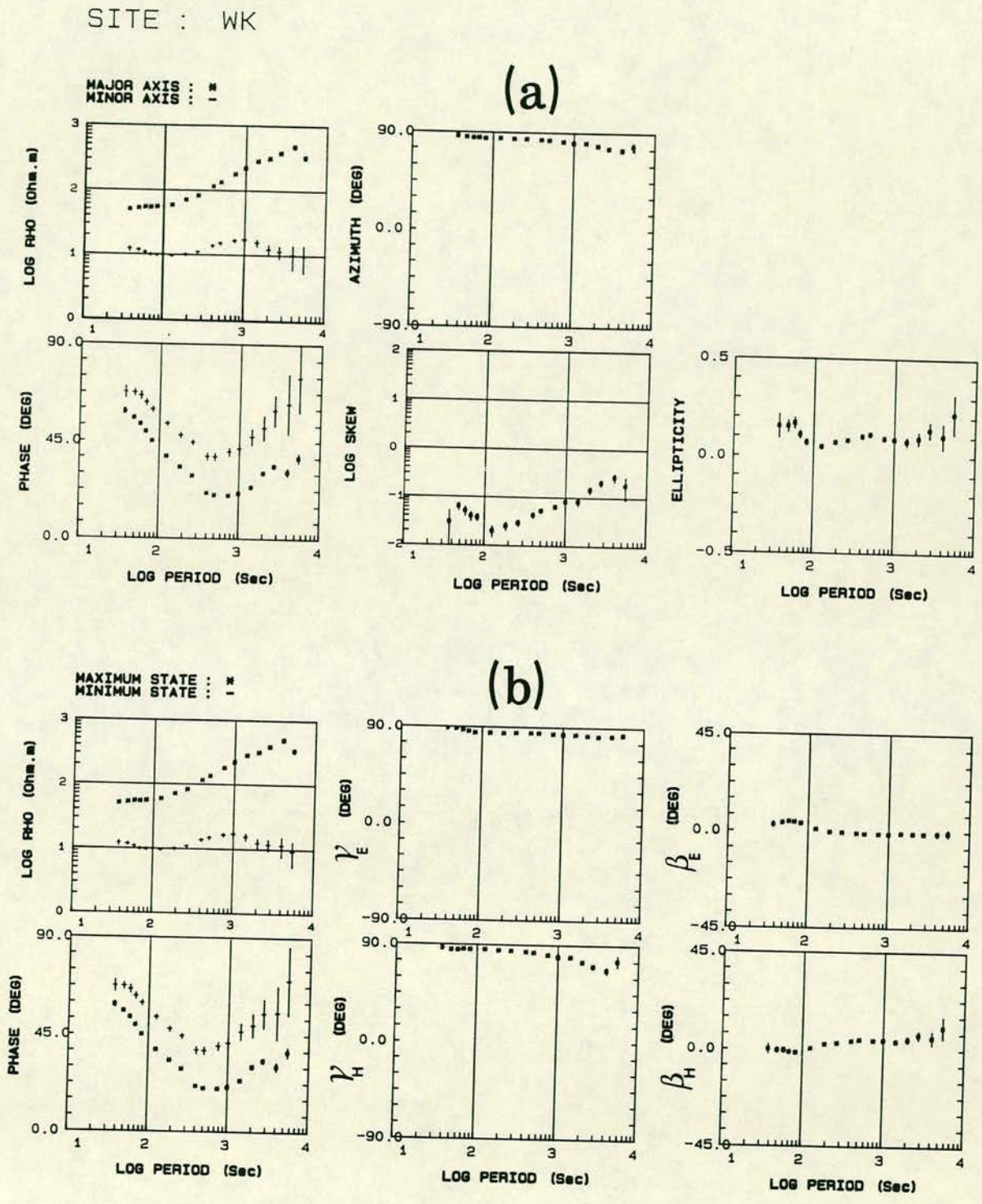


FIGURE 8 **(a)** The conventional analysis results for site WK.
(b) The UD results for site WK; the skew angle is not shown.

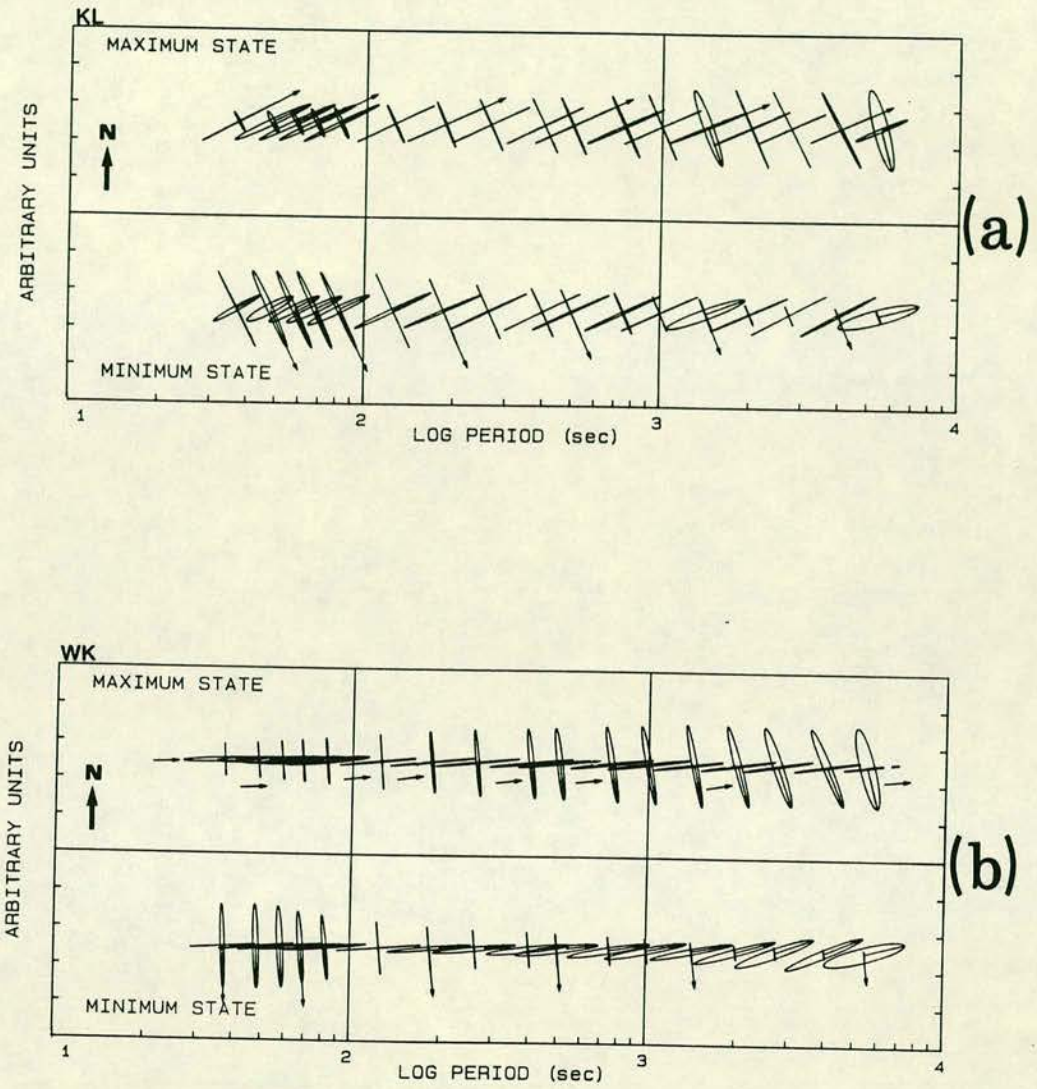


FIGURE 9 The 'state mode' of presentation for **(a)** site KL, and **(b)** site WK. The Figures show the characteristic states of the impedance tensor as per Figure 2a, and as a function of period (T). The response at a given T is *always* generated by a unit amplitude magnetic component, and normalized to the corresponding eigenvalue $\mu^{\pm}$ (electric field component) of the tensor. In this way, the relative amplitude of the electric and magnetic field components in a characteristic state, depicts their dynamic variation as a function of period. The direction of the electric field components is indicated by the arrows; this is the only function they serve and they must not be assigned any physical significance.

the evidence at hand, one cannot provide a clear explanation of the phenomenon. However, this may not be as awkward as it appears, for such behaviour may signify the influences of the regional current system, linking the North and the Irish seas. From Figure 5, we observe that WK is located over an area with a denser regional current stream function, and a steeper vertical (anomalous) magnetic field gradient (Figure 4), than the corresponding location of KL. It is therefore plausible that the regional current system exerts more influence on the E-states at WK and produces the observed linear polarization. The source of the elliptical polarization of the H-states may be sought in the gradient of the anomalous vertical magnetic field. The absence of appreciable vertical magnetic fields and intense current flow, may well be the explanation for the linear polarization observed in KL. One may also note that the WK maximum and minimum apparent resistivities are parallel for $T < 500$sec, and begin to diverge at periods > 500sec. KL displays a similar behaviour but its maximum and minimum apparent resistivities only begin to diverge for $T \geq 1000$sec. Also, at these long periods the apparent resistivities from both sites show a similar character. It has been shown (e.g. Banks and Beamish 1984), that at very long periods, the mode of induction depends on the large scale regional system. Therefore, I am tempted to assert that the cause of the divergence, at long periods, originates at depth, and its extent over the available bandwidth is controlled by the relative intensity and depth of the regional current system underneath the site of measurement.

IX. Site EB

The conventional analysis results from this site are shown in Figure

10a. Figure 10b shows the results of the UD. Figure 10c shows the comparison of apparent resistivities obtained by the two techiques. Finally, the Vertical Field Transfer Functions (VTF) for the same site are presented as a real and an imaginary part with their associated azimuths in Figure 10d. The VTF can be qualitatively explained by a conductivity distribution that varies from strongly 3-D to almost 2-D, and back to 3-D again, as a function of depth (frequency).

Attention is drawn to the shape of the β_H (and β_E) curves, and their similarity to the *real* part of the VTF. $\beta \rightarrow 0$ at the same frequencies where the real part of the VTF *maximizes* and the imaginary part minimizes. This shows that linearly polarized fields are associated with this part of the bandwidth. The conventional ellipticity conveys similar information. The conventional azimuth θ_c compares with γ_E; they are similar in shape, but they differ by approximately $10°(\theta_c > \gamma_E)$. Also note that the azimuth γ_H of the H-frame, is very different from θ_c. The shape of both γ angles resembles that of the imaginary part of the VTF. The conventional skew and ellipticity display curves of similar shape to γ_H and β_H, and apparently conveys similar information. However, their interpretation may be seriously misleading, as we shall see shortly. From Figure 10c, where the plots of the conventional analysis and the UD apparent resistivities are overlayed, it can be seen that the maximum and minimum resistivities returned by the two methods are not the same. The conventional major and minor axes actually lie somewhere in between the extremal values, as obtained by the UD; the nearly order of magnitude difference in the estimation of the extremal values is very impressive, and summarizes the need of an alternative rotation procedure. It is worthwhile noting that the conventional analysis does not achieve the extremal values even for

the linearly polarized part of the bandwidth. This is a very interesting feature that deserves more attention.

The UD analysis may allow inferences about the mode of induction in a particular location, since it conveys three-dimensional information. Figure 10e shows the 'state mode' for EB, as per VII.2.1. It becomes clear from the figure, that when $\beta \rightarrow$ min, the electric and magnetic characteristic states remain non-orthogonal. This is a very peculiar feature that reveals horizontal characteristic states, but certainly not a two-dimensional character. Furthermore, the maximum eigenvalue of the tensor is aligned with the linear feature of the current flow (Figure 5). However, one would expect the opposite to be true, i.e. the minimum eigenvalue to face the direction of current flow, since it defines the orientation of the 'most conducting path'. The azimuths γ_E and $\gamma_E+\pi/2$, of the maximum and minimum electric components for the period 750sec, are plotted in Figure 5; all azimuths are corrected for the declination of the magnetic field (N9°W). This period belongs to the linearly polarized portion of the bandwidth. The two features may provide information about the mode of induction in the vicinity of the site. The linear current feature is driven by regional induction processes that involve the Atlantic Ocean and the North and Irish seas (Banks et al, 1983). The site is located near a relative high of the anomalous vertical field and faces steep gradients towards the south and the west. Accordingly, the dense current stream function lines indicate a relatively intense current flow, that may be dominant over the whole bandwidth. Therefore, it may be possible that the impedance tensor at EB is biased by the regional system and does not convey local structural information. The main difference in the shapes of the E- and H-frame curves are that the former are flatter versions of the latter; this

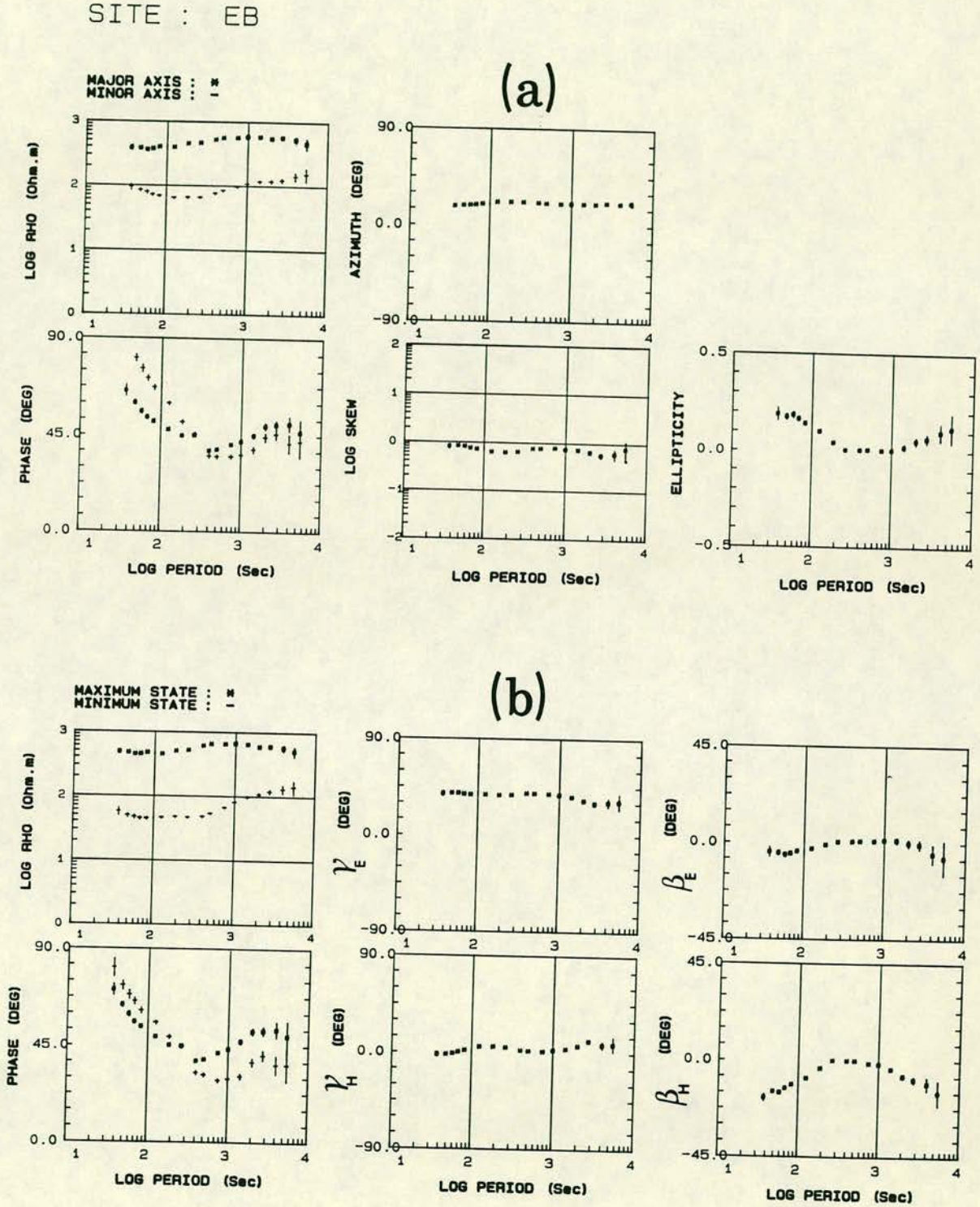
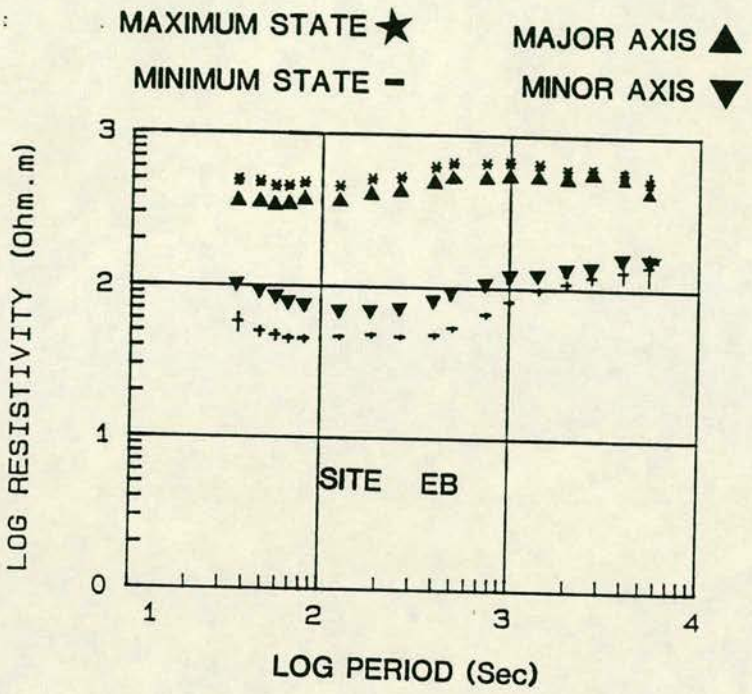
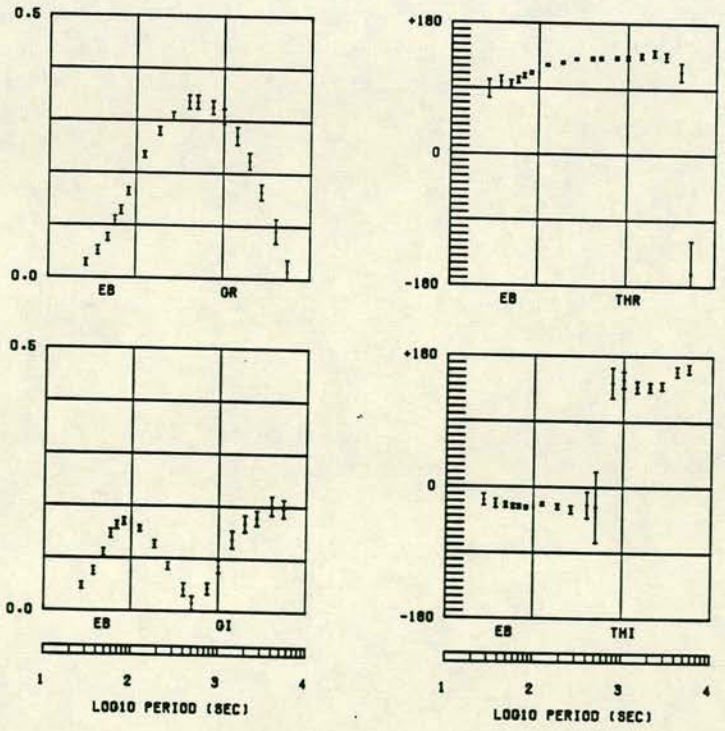


FIGURE 10 **(a)** The conventional analysis results for site EB.
(b) The UD results for site EB; the skew angle is not shown.


MAXIMUM STATE ★
MINIMUM STATE –
MAJOR AXIS ▲
MINOR AXIS ▼
LOG RESISTIVITY (Ohm.m)
3
2
1
0
SITE EB
1
2
3
4
LOG PERIOD (Sec)

(c)


0.5
0.0
EB
QR
+180
0
-180
EB
THR
0.5
0.0
EB
QI
+180
0
-180
EB
THI
1
2
3
4
LOG10 PERIOD (SEC)
1
2
3
4
LOG10 PERIOD (SEC)

(d)

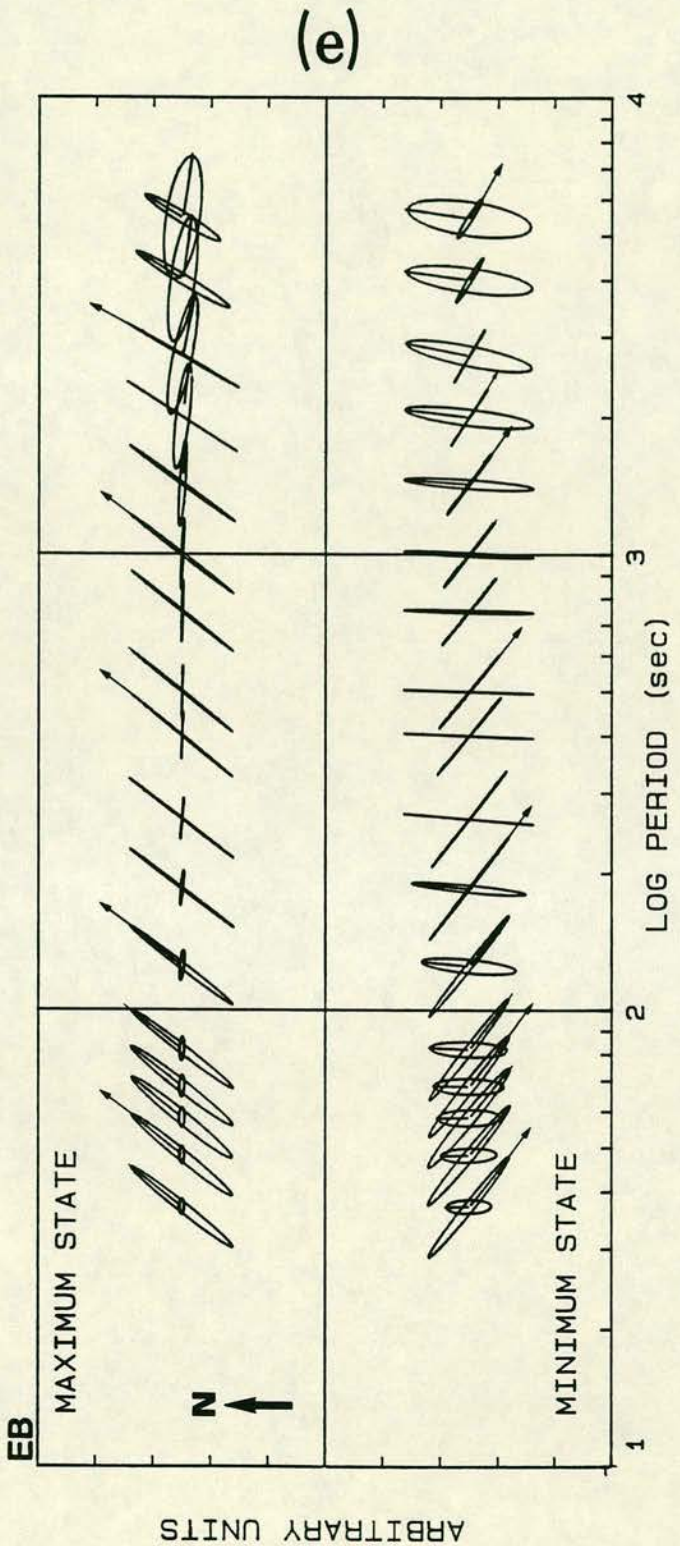


FIGURE 10 (e) The state mode as per Figure 9, for site EB.

shows that the behaviour of the E-states is more uniform than that of the H-states, for most of the bandwidth. If such an interpretation is correct, then the linearly polarized part of the bandwidth should coincide with the depths of the current flow, which is horizontal. Above the main current concentration there may exist perturbations due to the local structure and induction mode, that tilt the total E- and H-fields, thus generating elliptical polarization. For T > 1000sec, the behaviour of the tensor shows a character quite similar to that of WK, with linearly polarized electric, and elliptically polarized magnetic components.

X. Site PW.

Figure 11a displays the results of the conventional analysis, while Figure 11b those of the UD. Figure 11c is a direct comparison of apparent resistivities by the two procedures, and Figure 11d shows the VTF, as per site EB. Finally, the 'state mode' diagram is provided in Figure 11e. The MT data from this site allow some very interesting observations. Firstly, observe that for periods < 200sec, the azimuths γ_E and γ_H are stable, although the angles β_E and β_H indicate a substantial degree of three-dimensionality for this site. Above 200sec the situation changes dramatically, due to the onset of intense 3-D influences that cause the horizontal azimuths γ_E and γ_H to rotate; the gradient of these curves must be paired with the bulge observed in the curves of the vertical angles β_E and β_H, indicating the EM field tilting due to the 3-D effects. The maximum inflection of the β angle curves occurs at about 400sec; this period also marks the point of steepest gradients in the horizontal azimuths. Such effects should not be

unexpected because, assuming that the rotations of the horizontal azimuths are due to some 3-D geoelectrtic structure, the most powerful effects should occur nearby the areas of most rapid change in the structure. Finally, above 2000sec, the sharp changes ease and the horizontal azimuths appear to stabilize with respect to a new (regional) strike. The diminishing β angles also appear to indicate that the regional regime is one of reduced dimensionality. In short, the mean initial horizontal azimuths are γ_E = N42°W, rotating towards W, and γ_H = N22°E rotating towards E. The final horizontal azimuths are γ_E = N98°W and γ_H = N80°E.

The study of Beamish (1986b) indicates, that this site displays definite 3-D characteristics, and the present analysis supports this notion. The cause of the 3-D effects is not exactly known, but it is thought to be a complexity of geological features rather than a single body (Banks et al 1983; also see figure captions herein). Thus, the initial azimuths can be attributed to local structure, while the final (long period) azimuths may be due to large scale induction effects. The real part of the VTF is consistent with the concept of 3-D influences at the relatively short and medium periods, becoming increasingly dominated by strong 2-D processes at long periods. The location of the site over an area of weak vertical magnetic field gradients and current stream function at the period of 750sec (Figure 5), may indicate that local induction effects predominate at short and intermediate periods. This idea is strongly supported by the results of Banks and Beamish (1984), and Banks et al (1983); their analysis is based on the Figures 4 through 6, and indicates that the Vale of Eden, to the west of the site does not make conductive links, and therefore insulates the Alston Block from the Irish Sea. The current stream function over the Alston Block, is at

least partially shallower than 5Km for this period (Banks et al, 1983), and therefore, the resulting impedance ought to be strongly influenced by local induction effects. However, the azimuths at very long periods have an almost E-W orientation, indicating a transverse current flow across England, and the tensor appears to have been produced by linearly polarized fields. Thus, the flexure of the β angle curves may indicate a combination of gradual diminishing of local 3-D contributions, and the onset of the regional current influences, that alter the behaviour of the EM field and mark the transition from local 3-D to regional induction processes; due to their significant lateral extent, the latter can be satisfactorily be represented by a 2-D induction model.

The conventional azimuth conveys none of the aforementioned information. It becomes discontinuous at 750sec, which corresponds to the period where both horizontal azimuths of the E- and H-frames exceed 45°. The 'new' azimuth (above 750sec) is associated with a new 'major' and 'minor' axes, which is a veritable source of confusion when interpretation is required. Unfortunately, the conventional azimuth is meaningless at this site, especially when compared with the VTF. It is also quite obvious from Figure 11c that the conventional major and minor apparent resistivities lie somewhere between the actual extremal values. The conventional apparent resistivities and phases are discontinuous and shifted because of the limitations of the procedure; this is another well known problem that hinders any interpretation attempts when impedances like the one in site PW are obtained. Interestingly enough, the skew parameter conveys exactly the kind of information the horizontal azimuths γ_E and γ_H do. From its definition, it is quite easy to see how, however, the normal use of the parameter would never have

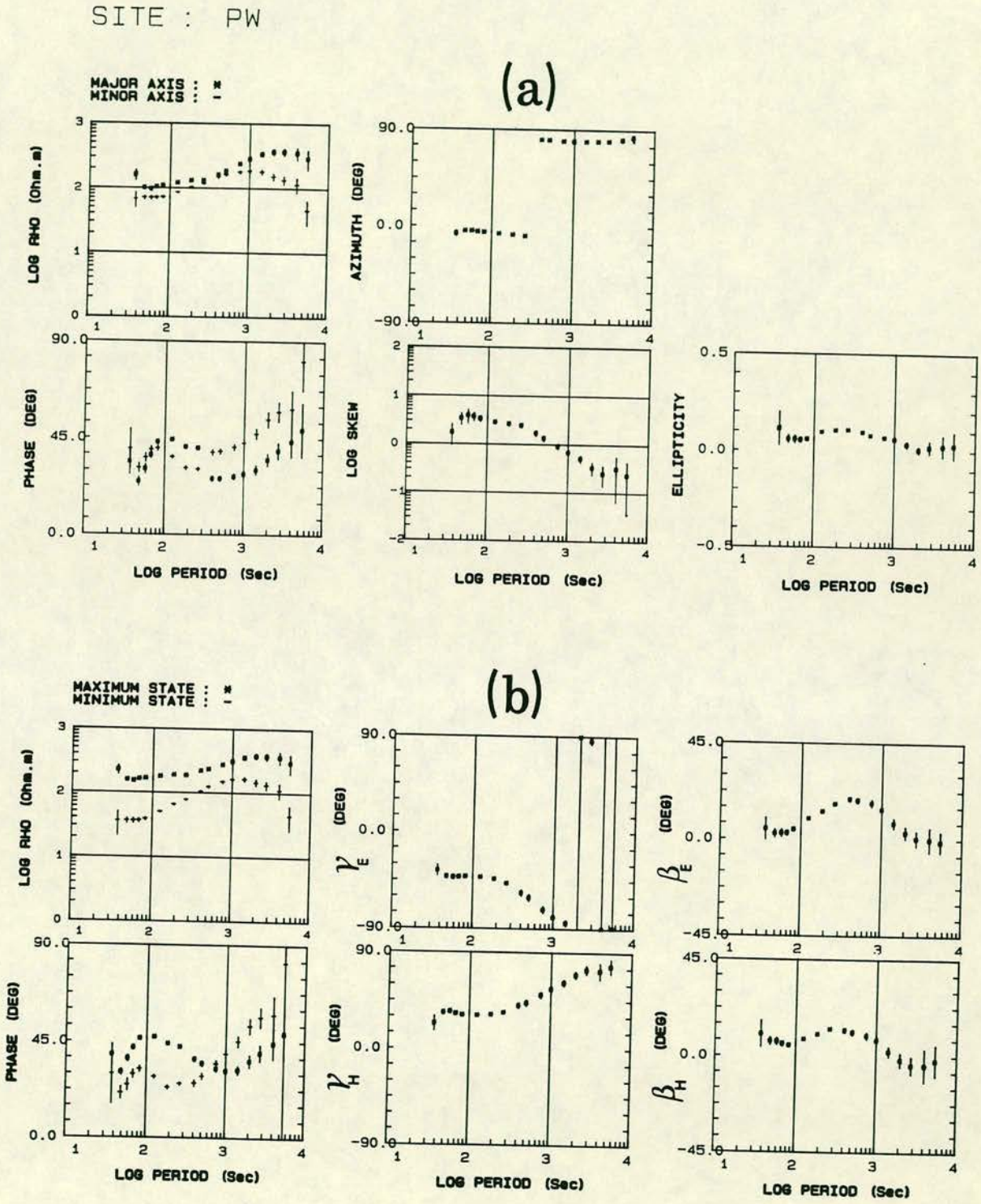
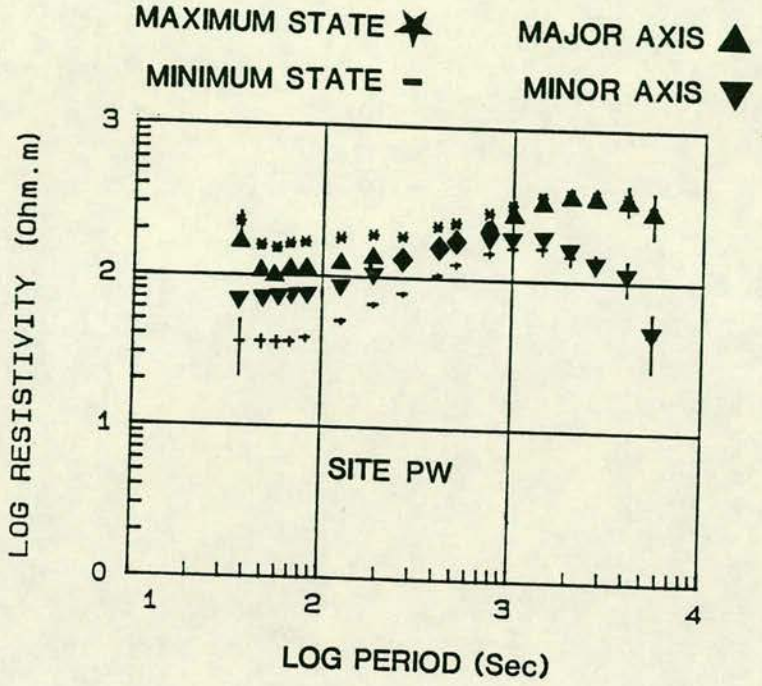
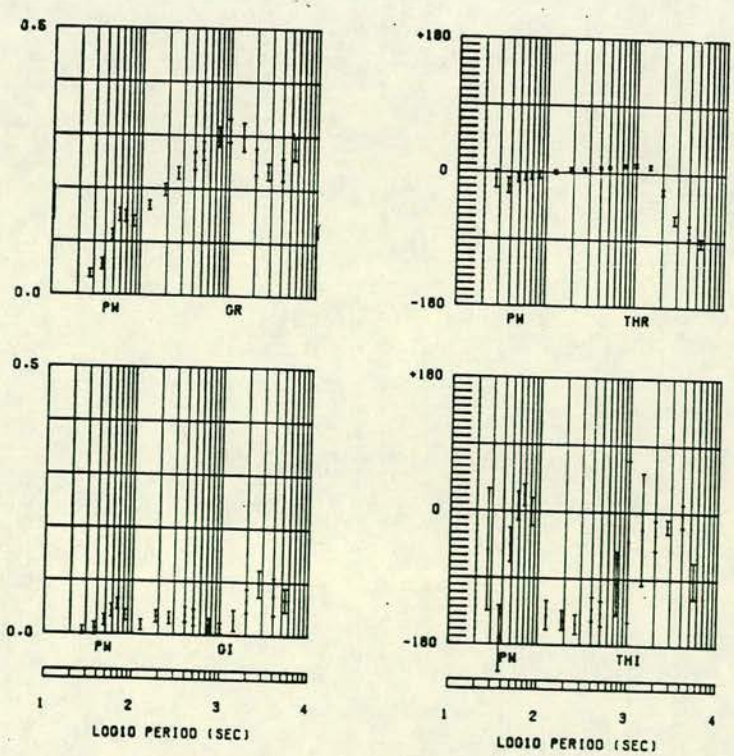


FIGURE 11 **(a)** The conventional analysis results for site PW.

(b) The UD results for site PW; the skew angle is not shown.


MAXIMUM STATE
MINIMUM STATE
MAJOR AXIS
MINOR AXIS
LOG RESISTIVITY (Ohm.m)
SITE PW
LOG PERIOD (Sec)

(c)


PW
GR
PW
THR
PW
QI
PW
THI
LOG10 PERIOD (SEC)
LOG10 PERIOD (SEC)

(d)

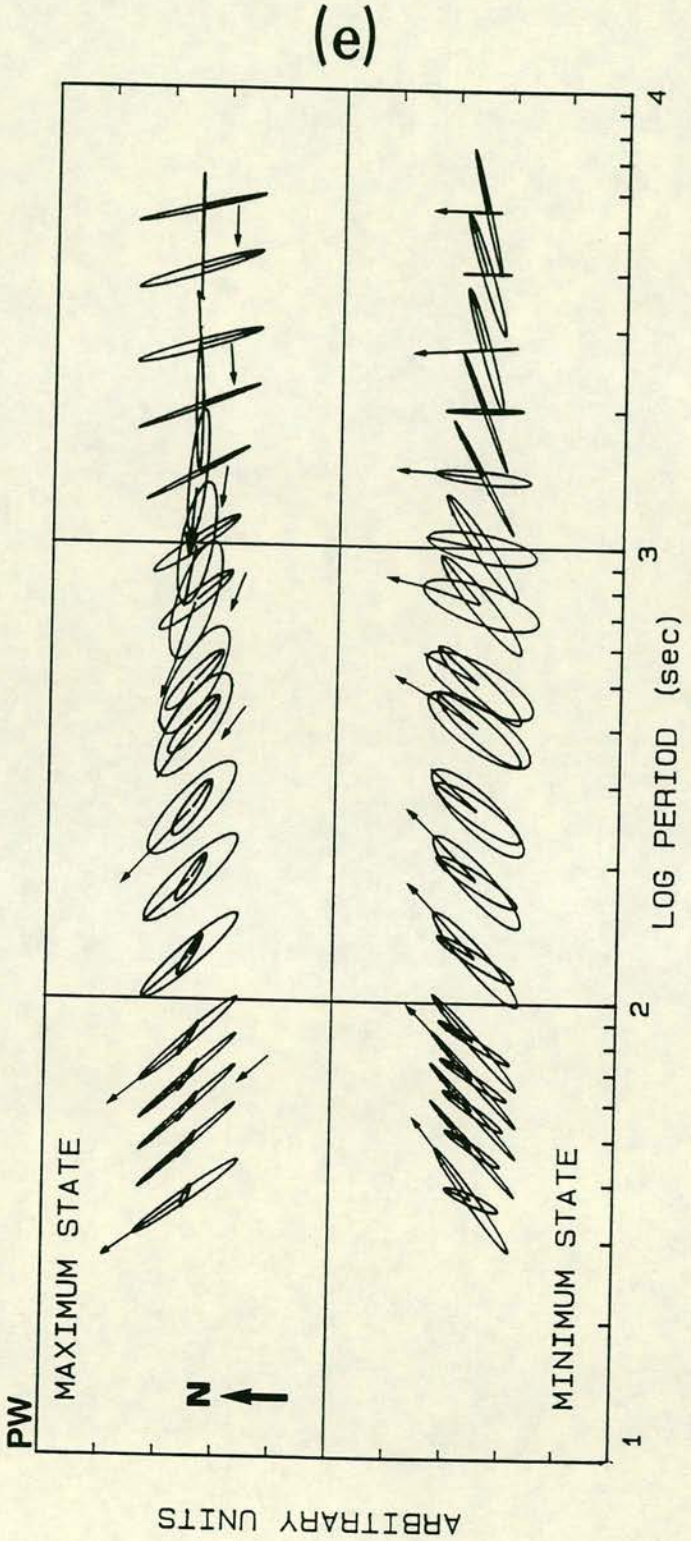


FIGURE 11 **(e)** The state mode as per Figure 9, for site PW.

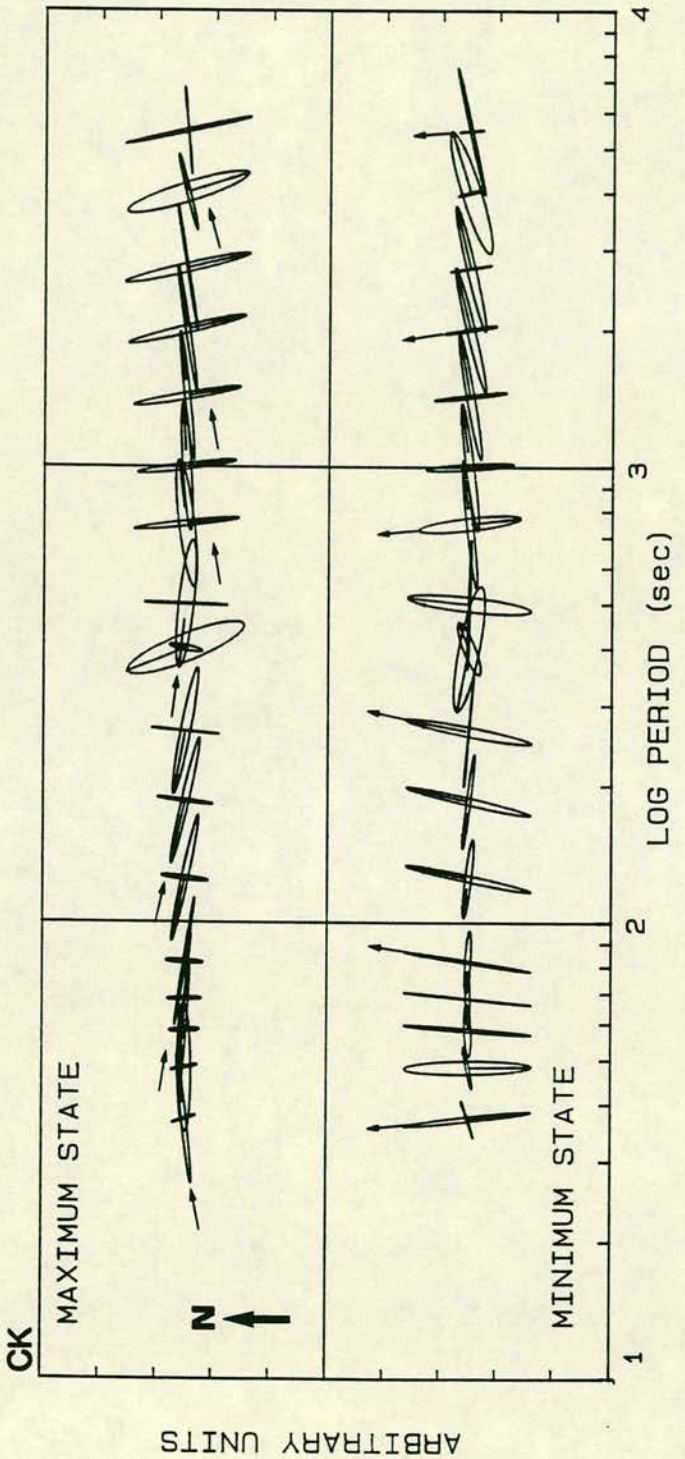


FIGURE 12 The state mode as per Figure 9, for site CK.

allowed inferences such as above to be made. Ellipticity has the right shape, but insufficient magnitude. In summary, the conventional analysis of PW can return totally useless results, in terms of our ability to correctly interpret the structure.

XI. Some observations on the mode of induction over S. Scotland and N. England.

Every informed person that may read the following few paragraphs would probably warn against making inferences based on such a limited data set. I would like to state that I am aware of the potential loopholes, and that I do not provide solutions, merely observations. An additional site to the four analysed above, is hereby introduced; this is Site CK, designated as Site A by Beamish (1986b), and as Site 2 by Beamish (1986a). The azimuths of the electric field states are plotted in Figure 5 for the period 750sec, while other information is provided in 'state mode' diagram only, in Figure 12.

The common feature of sites EB, WK, KL and CK is that at short ($T < 100$sec) periods they display, in general, elliptically polarized waves. At intermediate ($T < 1000$sec) periods they show a greater degree of linear polarization, while at long ($T > 1000$sec) periods the polarization becomes elliptical again. Elliptically polarized E-states exist for short periods, that, with the exception of CK, definitely become linearly polarized at intermediate and long periods; moreover, (again with the exception of CK), the polarization of the H-state does not display a common behaviour. The degree and extent of polarization, naturally varies from site to site; however all sites display very

similar polarization characteristics at long periods. Site CK is situated near the Eskdalemuir Geomagnetic Observatory; it shows elliptically polarized E-fields, and linearly polarized H-fields at intermediate periods, as opposed to EB and WK. It also shows a very clear rotation of the azimuth for T < 1000sec. The results of Banks and Beamish (1984) show that at long periods (750sec), the regional response to a horizontal unit amplitude northward or eastward H-field, is a very clear low in the amplitude of the anomalous vertical field, at the location of the site (Figure 4). Note also, that the regional current pattern is appreciably weakened and deflected to the east of the site. At short periods, the same study shows that at short periods, a northward or eastward unit amplitude horizontal H-field produces virtually no vertical field response around CK (e.g. Figure 6). The different character of CK with respect to these sites may be due to a resistive structure to the east of the site, that is corroborated by the evidence, and deflects the regional current stream function. Therefore, at short and intermediate periods, the site probably contains only local information, with the elliptical polarization of the E-fields being produced by the structure. Beamish (1986b) recovers acceptable D^+ solutions for T < 500sec, while his detailed inversion exercise (Beamish 1986a), reveals a very resistive upper-crustal layer (of the order of 10^5Ohm.m) underneath the site. The regional system would appear to predominate at very long periods. At 750sec, the behaviour of the tensor may be due to the combined influences of the large scale resistive structure and the regional induction mode. PW is also thought to convey local induction information at short and intermediate periods, as detailed in VII.2.3. I think it suffices to say that the results at PW show exactly the opposite character to those at EB,KL,WK save for the

long period band, while they show a degree of similarity with CK at short and intermediate periods, as far as the mode of polarization is concerned.

I believe that the common polarization features are far too similar to be coincidental for such widely separated locations, and possibly convey information about the regional mode of induction. I quote from Banks et al (1983): "Our results indicate that, in the period range $10-10^4$s, transfer functions for the north of England are influenced by the magnetic fields associated with three different modes of induced current flow. At periods greater than 2000s, they depend on the direct fields of currents flowing in the Atlantic Ocean and in the shallow seas around the British Isles. Between 300 and 1500s, the fields at many sites are produced by the perturbation of currents driven through the land as part of a thin-sheet induction process which includes the shallow seas. At periods < 100s, the transfer functions of sites which are sufficiently well insulated from conducting pathways between the seas detect currents produced by local induction in isolated regions of high conductivity". I think that the UD results from the five sites presented here, provide Magnetotelluric overtures to the results of the authors. EB appears to be influenced beyond recovery in terms of useful local information. CK and KL appear to be sufficiently insulated and are interpretable, with acceptable D^+ solutions for the period range < 500sec. WK is not, but still is not damaged beyond recovery; the regional current flow appears to follow an E-W path, and this corresponds with the geological structural trends at the vicinity of the site. Therefore, if a simple geoelectric structure also exists at depth, it may not be lost beyond recovery. At any rate, acceptable D^+ solutions exist for this site as well. However, Beamish (1986a,b) had to reject all determinations above

this limit in order to obtain physically meaningful results at the three sites. No 1-D interpretation is recoverable from PW. The common polarization characteristics from all sites for T > 1000sec can easily be appreciated if they are due to some common cause, i.e. the regional induction system. I cannot, at present, propose an explanation for the mechanism that would generate the tilt of the magnetic field; however, I feel that these observations require more attention. The azimuths of the long period states are, of course, site dependent, but PW,WK and CK show good agreement. One would expect all azimuths to adapt to the 'local' perturbations of the regional mode of induction, and they compare well with the vertical field results available for the 750sec period (Fig. 5).

In conclusion, I think that the findings of the UD may prove to be consistent with the results from the large scale GDS experiments conducted by the above authors. If that is so, they provide a clear indication of the degree of care that must be exercised while interpreting MT data aquired in the studied areas, or indeed any similarly situated region. It would appear that little meaningful information about 'local' induction is available at long periods because results may be greatly biased by the regional induction mode. Intermediate periods may also be influenced; site EB possibly provides a good example of the degree of bias that can be experienced at a single site, while similar assertions can be made for CK and WK. Therefore, a great deal of caution must be exercised, and certainly, more information than that 'conventionally' obtained must be made available, before interpretation is attempted. I believe that the UD provides a wealth of quantitative interpretational aids, including information about the vertical components, that may prove invaluable in difficult situations.

CHAPTER 2

PART 1:

A DISCUSSION ON THE NATURE OF THE MT RESPONSE FUNCTION AND SOUNDING CURVE

The following discussion is a survey of some simple MT relationships that are just too easy to overlook! They are easy to overlook because they arise from a consideration of the MT response functions in the time domain, with which the MT research community is not very familiar. In the time domain, we have to deal with waves (and wavelets) more explicitly than in the frequency domain, where the bulk of the research is being carried out; the frequency domain formulation of the MT problem, somehow facilitates the downrating of the wave nature of passive induction processes, as well as their time component, and the overrating of their 'diffusion' component. These simple facts (waves) are frequently mentioned, but I doubt whether they are significantly appreciated in the interpretation of MT problems and data. The following few pages contain an attempt to understand some aspects of the time domain component in the MT problem. I must state that a large part of the discussion does not contain something that has not been said before, at least outside the MT community, or does not exist, in one form or another, in well known MT literature. This work attempts to fit a few (hopefully) useful time domain concepts and notions into the framework of the MT problem, and succeeds only to 'scratch' the surface of a very interesting and fascinating subject. However, many time domain concepts will be understood to fit conveniently into the 'sophisticated' theoretical formulation of Weidelt (1972) and Parker (1980). No solutions to any problem are on offer; merely an

unconventional review of some, I believe, basic properties of the MT response function.

I. The Earth response function as a finite order polynomial

In the following discussion I shall only be interested in the case where only a finite portion of the EM field spectrum and/or only finite data realizations are available. In most practical situations these two conditions become equivalent (e.g. band-limited digital data). This means that only finite order realizations of the physical systems under study are attainable, and frequency transformation can only be carried out by using approximations of the Fourier Transform (FT). In order to study the properties of such systems it is more convenient (and correct) to use the z-transform instead of the Fourier transform to obtain the frequency response of a physical system. The z-transform of x(t), being equivalently called the frequency transform of x(t), is defined through the polynomial

$$X(z) = \sum_{t=0}^{N} x_t \, z^t$$

with $z=\exp(i\omega)$ being the conventional definition of the z-transform. A limiting operation may convert the z-transform into a Fourier integral when its order increases. In general, the z-transforms and Fourier transforms are related by

$$z = \exp(i\omega), \quad i\omega = \ln z, \quad \omega = 2\pi n - i\ln z \qquad (2)$$

As can be seen in (2), the mapping between the two planes is non-linear. Thus, simple ratios of polynomials in z will not transform into simple ratios of polynomials in $i\omega$, and vice versa. However, the existence of the relationship indicates that for any analysis that can be worked out in one plane, there exists an equivalent analysis that can be worked out

in the other.

The study of the MT physical system is based on two fundamental assumptions: **(a)** The dependence of the electric field (E-field) on the magnetic field (H-field) is linear, and, **(b)** at any time t_0 the value of the H-field depends on past and present values of the E-field, and vice versa. Physically this is equivalent to requiring that the system be passive and stable, i.e. no response is observable prior to its excitation by an external source. This means that the MT physical system comprises a *causal* system. Conditions (a) and (b) imply that at a given instant t_0 the system has only memory, and no anticipation components, i.e. it is equivalent to an AutoRegressive (AR) process. We write, assuming a one-dimensional (1D) dependence for simplicity,

$$h(t) = y_0 e(t) + y_1 e(t-1) + y_2 e(t-2) + \ldots + y_N e(t-N)$$

This is the convolution operation

$$h(t) = e(t) * y(t)$$

$$y(t) = 0, \quad t < 0$$

By taking the z-transforms of both sides,

$$H(z) = E(z)\, Y(z)$$

from which we obtain

$$E(z)/H(z) = 1/Y(z) = 1/(y_0 + y_1 z + y_2 z^2 + \ldots + y_N z^N) \qquad (3)$$

and we agree that the z-transforms of the EM field components, i.e. the EM field spectra are defined by the z-transforms of their autocovariance (AC) functions or any equivalent AR representation (e.g. Burg 1975, Ulrych and Bishop 1975, and others). The transfer function $1/Y(z)$ represents an all-pole (AR) time invariant system, comprising an inverse polynomial in z, of degree N. $Y(z)$ is defined as the *admittance* function of the MT system. The inverse of the function $Y(z)$ must be an *invertible* function $Z(z)$, such that

$$Y(z)\ Z(z) = 1$$

Then, we shall call Z(z) the *impedance* function of the MT system. This function would be described, in the time domain, by the convolution

$$e(t) = h(t) * z(t)$$

Invertibility of the two functions is a condition that is necessary and sufficient for physical realizability, i.e. for the very existence of the two way causal connection between the E- and H- fields. In turn, physical realizability is a necessary and sufficient condition for the stability of the MT physical system, a concept that is implied by the requirement that the energy contained in the passive system be finite over any time window, since no energy sources or sinks are supposed to exist in the system. The mathematical condition for stability (and two way invertibility) is that no roots of either Z(z) or Y(z) are allowed to lie inside the unit circle (e.g. Robinson 1962, 1967), i.e. $|z_j| > 1$ where z_j refers to a root of either Z(z), or Y(z).

We see, therefore, that the physical conditions imposed on our problem produce the following properties on our transfer functions:

(a) The transfer functions are one-sided functions of time, one property that is identified with causality.

(b) The transfer functions are stable two-way invertible polynomials, of degree N, with at least N roots $|z_j| > 1$. By the fundamental theorem of algebra, they can be factorized in terms of binomial factors, as

$$Y(z) = y_N\ (z - z_1)\ (z - z_2)\ (z - z_3)....(z - z_N) \qquad (4)$$

The multiplicity of the roots in (4) is taken to be one (simple roots), a concept that will be clearly understood later.

(c) Due to their polynomial nature, the transfer functions can be evaluated at *any* z on the unit circle, i.e. they possess an absolutely continuous frequency response.

(d) Transformation of (4) back into the time domain yields

$$(y_0, y_1, \cdots y_N) = y_N\ (-z_1,\ 1)*(-z_2,\ 1)*\ldots*(-z_N,\ 1) \qquad (5)$$

i.e. the time function associated with the $(z - z_j)$ factor is $(-z_j, 1)$. Thus, the transfer functions result from a cascade convolution of N two-coefficient (elementary) operators. Under the condition $|z_j| > 1$ this yields

$$|y_0| > |y_1| > |y_2| > \ldots.$$

Time domain operators exhibiting such properties were called by E. Robinson Minimum Delay (MD) operators. Their frequency domain equivalent property is that of Minimum Phase (MP) (e.g. Claerbout, 1976). The physical meaning of the MD/MP property is that the transfer of energy between the EM field components is squeezed up as close to time t=0 as possible, following the excitation of the system, (i.e. it is loaded in the front end of the system). This, in fact, is a physical consequence of the causal connection between the EM field components, and determines that the minimum amount of dispersion occurs during the transfer. Therefore, (5) is a form of Robinson's energy delay theorem (Robinson 1962). For an electrically layered Earth, a delay in the transfer of energy is introduced due to the conductivity structure, and the transfer function assumes the form of a finite order MD wavelet, conveying information about this structure. This however shall be the matter of a later more detailed discussion.

(e) For finite data realizations, the transfer functions contain finite energy, in the sense that they describe a finite amount of energy transfer between the EM field components of a passive system, and, therefore, they must show positive energy dissipation. In fact, the requirement is that they show positive energy dissipation over any time window from minus infinity up to any time t preceding or including the

present. It can be worked out, (e.g. Claerbout 1976), that such properties necessitate that the transfer fuctions be *positive real*. This means that the polynomial coeficients and their cosine transform must be positive. As a result, the following properties are equivalent: The real parts of their frequency transforms are even functions of frequency, and the imaginary parts are odd functions of frequency. The real and imaginary parts of their frequency transforms are Hilbert transforms of each other. Although theoretically expected, the positive real property is not always numerically retrievable from actual data, even in the low noise case. The MD behaviour and the positive cosine transform however, are properties that must be strictly obeyed by a physical response function, and unequivocal indicators for its validity.

It is well known that any physical spectrum admits a MD spectral factorization of the type (4) and/or (5) (e.g. Robinson 1967, 1980). Therefore, the power spectrum of E(z) and H(z) can be approximated by a finite order cascade convolution of MD dipoles; this comprises the treatment of their AC function as a finite degree polynomial (e.g. Robinson 1980), and is closely related to their minimum delay (Maximum Entropy) AR representation (e.g Burg 1975, Jaynes 1982). Thus, from (3) the transfer functions are defined both as rational functions, in terms of the poles and zeros of the minimum delay factorizable spectra E(z) and H(z) and in terms of their inverse minimum delay operators. In either case, there exist points in the z-plane where Z(z) is non-analytic, i.e. it is meromorphic. Because of (3), the transfer functions admit, by definition, the partial fraction expansion

$$Z(z) = \sum_{j=1}^{N} \frac{u_j}{(z - z_j)} \qquad (6a)$$

in terms of the singularities of Y(z). In order to define the constants u_j, we multiply (6a) by $(z-z_j)$ and take the limit as $z \to z_j$. Thus, all the terms of the partial fraction expansion vanish, except the one containing the term $(z-z_j)$ in the denominator. Then,

$$u_j = \lim_{z \to z_j} [(z-z_j)\, Z(z)] = \lim_{z \to z_j} [(z-z_j)/Y(z)]$$

This equation is indeterminate since the factor $(z-z_j)$ appears in the numerator and the denominator. We evaluate the limit using de l'Hospital's rule to obtain

$$u_j = 1 \,/\, Y'(z_j) \qquad (6b)$$

where $Y'(z_j)$ denotes the derivative of Y(z) at z_j. By (6b), the constant u_j is defined as the *residue* of Z(z) at z_j. Thus we see that the impedance function Z(z), and equivalently the admittance function Y(z), are given in terms of a function u_j that is constant everywhere, except for N finite points of discontinuity. The function u_j compares with the spectral function $\alpha(\lambda)$ utilized by Weidelt (1972), in its finite form (Parker 1980, and others). However, it is not identical. It is almost a spectral function, firstly because it is a function related to $Z(\omega)$, and not to the complex response

$$C(\omega) = Z(\omega) \,/\, -i\omega\mu_0$$

discussed by the authors, and secondly, because it is defined in the z-plane, and is, in general, a complex constant. It must be noted however, that a necessary condition for a positive real function is that all the roots that exist on the imaginary-ω axis must have a real positive residue (e.g. Ghausi and Kelly, 1968). Due to the definition of the z-transform by (2), we can see that for *real* frequencies ω, the real positive representation of Z(z) in the ω-plane, will have all its roots on the $-i\omega$ axis, and therefore, we anticipate that $u_j \to a_j(\lambda) > 0$, where $a_j(\lambda)$ is the ω-plane spectral function of the

positive real wavelet z(t). The mapping of (6a) into the ω-plane has the equivalent form

$$Z(w) = \sum_{j=1}^{N} \frac{a_j(\lambda)}{(\lambda_j - i\omega)}, \qquad a(\lambda) > 0$$

which, within a factor of $1/-i\omega\mu_0$, is the finite version of Weidelt's (1972) expression for $C(\omega)$, which is inverted by Parker (1980) and Parker and Whaler (1981). Thus, we can construct analytic expressions for Earth response functions, on the basis of filter theory and energy dissipation considerations only.

II. The relationship of the polynomial Earth response to the geoelectric structure.

Let me now discuss how the zeros of the polynomial transfer functions are related to the resistivity structure within the Earth. Let us temporarily return to the ω-plane; then, let (d) be the depth variable, taken as posistive downwards. By neglecting displacement currents and assuming an $\exp(i\omega t)$ harmonic time dependence, the electric and magnetic fields are related by:

$$E'(d,\omega) = -i\omega\mu_0 H(d,\omega) \tag{7a}$$

$$H'(d,\omega) = \sigma(d)E(d,\omega) \tag{7b}$$

where the prime denotes differentiation with respect to d. Upon eliminating H, (7) produces

$$E''(d,\omega) = -k^2 E(d,\omega), \tag{8a}$$

$$H''(d,\omega) = -k^2 H(d,\omega), \tag{8b}$$

with $k^2 = i\omega\mu_0\sigma(d)$ denoting the wavenumber. The solutions to (8) must represent an exponentially attenuated wave propagating into the Earth, and out of it. For equation (8a) these are given by

$$E(d,\omega) = A \exp(kd) + B \exp(-kd) \qquad (9a)$$

(7),(8a) and (9a) produce

$$H(d,\omega) = c[A \exp(-kd) - B \exp(-kd)] \qquad (9b)$$

where $c = k/i\omega\mu_0$, and the amplitude factors A and B to be determined. The response function $Y(\omega)$ is given by $Y(\omega) = H(\omega)/E(\omega)$; differentiate with respect to depth, and use the quotient rule to obtain the first order non-linear Riccati equations

$$\partial Y(\omega)/ \partial d - i\omega\mu_0 Y^2(\omega) = \sigma(d).$$

Likewise, $Z(\omega)$ will provide

$$\partial Z(\omega)/ \partial d - \sigma(d)Z^2(\omega) = i\omega\mu_0.$$

In the case of a conducting half space, $\sigma(d)$ is a constant. Thus, for a given angular frequency ω, the response functions, and therefore the ratios of the EM field components will also be constant, so that $Y(\omega) = c$, and $Z(\omega) = 1/c$. The Riccati equations show that the response functions, and therefore the ratios of the EM field components will deviate from a constant, if and only if the conductivity distribution varies as a function of depth. For the layered Earth model considered here, the conductivity varies with finite jumps, at the layer interfaces. This implies that in the layered Earth model, the response observable at the surface is generated at the layer interfaces *only*, and at no other points. From (3), (and the equivalent ω-plane representation), we see that the zeros of $Z(\omega)$ correspond to the zeros of $E(\omega)$ one for one, i.e. they correspond to the break points of the discontinuous part of the spectral function of $E(\omega)$. However, in our passive medium, $E(\omega)$ does not exist before its cause, and the same applies to its spectral function. $E(\omega)$ is generated as a response due to some conductivity discontinuities, therefore, the zeros of $E(\omega)$, and by conjecture the zeros of $Z(\omega)$ should reflect these discontinuities. In my opinion, these simple relationships are not clearly stated in most

MT reference texts, or, rather, the usual approach is to treat the electrically discontinuous Earth with continuous mathematics (e.g. equations 9). The method is, naturally, correct, but the interpretation may not provide the full picture of the processes involved.

In the following, I will construct an expression for the zeros of the finite polynomial C(z), that are a function of the discontinuities of an electrically layered Earth structure, consisting of a finite stack of uniform conducting slabs. The derivation will be carried out in terms of standard wave propagation theory. The upgoing (U) and downgoing (D) waves, indicated in (9), propagate with a *phase velocity* given by

$$v = \omega \,/\, \mathrm{Re}(k), \quad \text{with } \mathrm{Re}(k) = \sqrt{|\omega|\mu_0\sigma(d)/2}$$

The velocity of propagation is thus a function of the medium. At this point I consider it helpful to point out that the phase velocity of a wave is a *real* quantity. This implies that we can actually construct and use a perfectly valid theory of a response function, by considering the behaviour of the *phase field* only. The phase field experiences no energy losses, but only time stretching that depends on the medium of propagation. In this way, one can avoid the considerable problems that arise if the attenuation of the source fields is taken into account; this would introduce a complex wave velocity, in the same way a complex wavenumber is introduced in (8). Thus, there are two ways to look into a problem that involves wave propagation. Reflection seismologists implicitly use the phase field, in the time domain treatment of their physical problem; Magnetotellurists prefer the frequency domain, and the repercussions implied by considering a pure diffusion process. The opinion of this author is that we must also look the other way, and investigate it rigorously.

At an interface between two media j and j+1, with different propagation properties, the waves are reflected and transmitted according to Snell's law of refraction.

The attenuation constant inside the j'th layer is

$$Im(k) = \sqrt{|\omega|\mu_0\sigma(d_j)/\ 2}$$

The reflection coefficient is given by

$$r_j = (\sqrt{\sigma_{j+1}} - \sqrt{\sigma_j}\ ) \ / \ (\sqrt{\sigma_{j+1}} + \sqrt{\sigma_j}\ ) \qquad (10)$$

The transmission coefficient is given by

$$t_j = 1 + r_j$$

The form of the reflection coefficient given above, is only one of several alternative expressions that can be worked out. For instance, we can readily write the reflection coefficient as

$$r_j = (Z_{j+1} - Z_j) \ / \ (Z_{j+1} + Z_j)$$

A detailed derivation of (10a) in terms of EM field properties can be found in Ward et al, 1973. From a purely wave propagation point of view, and assuming a unit amplitude magnetic field vertically incident at the j'th electrical interface, between the media with admittances Y_j and Y_{j+1}, conservation of energy at the interface demands that

$$Y_j\ 1^2 = Y_j\ r^2 + Y_{j+1}\ t^2$$

from which it can be worked out that

$$r_j = (Y_j - Y_{j+1}) \ / \ (Y_j + Y_{j+1})$$

Yet another expression for the reflection coefficient can be derived using the expression for the phase velocity; thus,

$$r_j = -(v_{j+1} - v_j) \ / \ (v_{j+1} + v_j)$$

At the j'th interface, the U- and D- going electric waves will be related as

$$\begin{bmatrix} D \\ U \end{bmatrix}_{j+1} = \frac{1}{t_j} \begin{bmatrix} 1 & -r_j \\ -r_j & 1 \end{bmatrix} \begin{bmatrix} D \\ U \end{bmatrix}_j$$

where the reflection coefficient r_j is taken from (10).

For the sake of simplicity, we will assume that the traveltime of the upgoing and downgoing electric waves through each layer is constant, and equal to that of the data sampling interval Δt. This implies that the two-way traveltime of the wave in a single layer will be given by $z = \exp(i\omega)$.

The assumption of equal traveltime introduces some sweeping conclusions about the behaviour of such a system. The time taken by the wave to traverse a single layer is given by $\tau = h/v$, where h is the thickness of the layer. Then,

$$\tau_1 = \tau_2 \rightarrow h_1/v_1 = h_2/v_2$$

By making use of the relationship for the phase velocity we obtain, after some elementary operations, $\sigma_1 h_1^2 = \sigma_2 h_2^2$, and by extrapolation,

$$\sigma_j h_j^2 = \nu^2 ,$$

a constant. This relationship will hold throughout the system, and implies that the attenuation experienced by the wave is equal everywhere. Because the quantity $\mu_0 \nu^2$ has units of time, the delay of the response arriving at the surface is also constant, and equal to the data sampling interval. Thus, the upgoing wave (response) from the top of the j'th layer will be delayed by $\exp[i\omega(j-1)/2]$, i.e. the time it took to traverse the j-1 overlaying strata. It follows immediately that the single layer two-way traveltime, taken to be $z = \exp(i\omega)$, should actually read as $z = s = \exp(kd)$ for transient waveforms, and $z = s' = \exp(i\omega \pm kd)$ for harmonically dependent

waveforms, where kd = $\sqrt{i\omega\mu_0\nu}$, i.e. depends only on frequency. Propagation media displaying such properties are known as *Goupillaud* type media and have been extensively investigated by reflection seismologists. The relationship of the concept of equal traveltime to reality, will be discussed at a later stage.

Using this assumption, the matrix relationship connecting the U- and D-going waves at the interface, can be augmented to include the delay experienced by the waves, as

$$\begin{bmatrix} D \\ U \end{bmatrix}_{j+1} = \frac{z^{-1/2}}{t} \begin{bmatrix} z & -r_j \\ -zr_j & 1 \end{bmatrix} \begin{bmatrix} D \\ U \end{bmatrix}_j$$

This relationship is nothing new; it can be found in every standard textbook on reflection seismology and geophysical signal analysis (e.g. Robinson 1967, Claerbout 1976), and is fundamental to all physical problems involving wave propagation in stratified media. Based on the above matrix relationship, define the polynomials

$$\begin{bmatrix} F_{j+1}(\frac{1}{z}) & G_{j+1}(\frac{1}{z}) \\ G_{j+1}(z) & F_{j+1}(z) \end{bmatrix} = \frac{\sqrt{z^{-j}}}{\prod_j t_j} \begin{bmatrix} z & -r_j \\ -zr_j & 1 \end{bmatrix} \begin{bmatrix} z & -r_{j-1} \\ -zr_{j-1} & 1 \end{bmatrix} \cdots \begin{bmatrix} z & -r_1 \\ -zr_1 & 1 \end{bmatrix} \quad (11)$$

that are related as

$$G_{j+1}(z) = z^j \, G_j(z^{-1}), \qquad F_{j+1}(z) = z^j \, F_j(z^{-1})$$

It is apparent that these polynomials describe the flow of energy through the system, as a function of cascaded reflections and transmissions of the waves due to the electrical interfaces. The total energy flow between two adjacent layers (i.e. the polynomials of degree j and j+1) is given by the Robinson (1967) recursion

$$\begin{bmatrix} F_{j+1}(z^{-1}) & G_{j+1}(z^{-1}) \\ G_{j+1}(z) & F_{j+1}(z) \end{bmatrix} = \frac{z^{-1/2}}{t_j} \begin{bmatrix} z & -r_j \\ -zr_j & 1 \end{bmatrix} \begin{bmatrix} F_j(z^{-1}) & G_j(z^{-1}) \\ G_j(z) & F_j(z) \end{bmatrix}$$

This recursion can be used to relate the waves observed at the surface to the waves observed at the j+1 layer as

$$\begin{bmatrix} D(z) \\ U(z) \end{bmatrix}_{j+1} = \frac{\sqrt{z^{-j}}}{\prod_j t_j} \begin{bmatrix} F_j(z^{-1}) & G_j(z^{-1}) \\ G_j(z) & F_j(z) \end{bmatrix} \begin{bmatrix} D(z) \\ U(z) \end{bmatrix}_0$$

If we assume that the index j+1 corresponds to an infinitely thick conducting half-space at the bottom of the electrically layered system, then there exists no upgoing wave coming from that region. Set $U_{j+1}(z) = 0$. In order to obtain an analytic expression for the waves (fields) observed at the surface, one must carefully consider the conditions at the surface. The atmosphere can be considered as a non-conductor, a medium in which EM waves can propagate. Thus, part of the upgoing wave will be radiated into the atmosphere, where it will propagate unattenuated, as a pure displacement (non-free surface conditions). The downgoing wave at the surface, can be identified with the excitation source. Although we generally consider harmonically dependent sources, for the limited scope of the present we can assume that the source is a unit spike (impulse) at time 0. Alternatively a unit amplitude plane polarized source described by a pure cos(ωt) dependence, as assumed above, is appropriate. In addition, by inspection of (11) one can see that $G_j(z) = -r_1 z + \ldots -r_j z^j$, i.e. it is a purely negative function. Under these circumstances one concludes that

$$U(z,0) \;/\; D(z,0) = F_j(z) \;/\; G_j(z) \qquad (12)$$

The response function observed at the surface will be given as a rational function, whose number of zeros corresponds to the number of conductivity discontinuities one for one. (12) can be expanded in partial fractions as in (6a). Then, the numerators u_j of the expansion will correspond to N finite points of discontinuity that correspond to the discontinuities of the conductivity profile, and will otherwise be constant.

The coefficients of the polynomial function $F_j(z) \;/\; G_j(z)$ are

equivalent to Kunetz's (1972) 'electric images' of the structure. Janoth (1984) uses the term 'Reflexionsantwort' which may be liberally translated as 'reflection electrogram/magnetogram'. Notably, Janoth's analysis turns out to be a reappraisal of Kunetz's original ideas, with quite interesting results. However, (12) differs from the response functions of the above authors in that they have considered a free surface boundary condition, i.e. one that totally reflects the upgoing wave back into the Earth. In conclusion, we can say that if the Earth structure consists of exactly N such discontinuities, the degree of the polynomial Z(z) sufficient to describe it is exactly N. Conversely, if in any numerical implementation for extracting z(t) we force the order of the wavelet to be $N_1 > N$, then N and only N of the filter coefficients returned will be significant, with the rest $N_1 - N$ coefficients tending to vanish. At least this is what theory says; however, the real life conditions are not so rosy. The above result has been obtained for a Goupillaud type model, assuming equal traveltime in each layer of the system. How physical is this assumption?

III. Is the one dimensional Earth a Goupillaud type conducting medium?

For discrete (and/or) digitally recorded EM field components, sampled at an equal time interval Δt, the linear operator y(t) can only be approximated as a discrete minimum delay wavelet $y_n = (y_1, y_2, \ldots, y_N)$, with coefficients y_j spaced Δt apart, and a polynomial z-transform

$$Y(z) = y_0 + y_1 z + y_2 z^2 + \ldots + y_N z^N$$

We are constrained by the requirement that *all* the roots of Y(z) be distinct, for physical realizability. If we agree that these correspond to conductivity discontinuities, it follows immediately that the latter

are revealed in cascade succession, at a time interval Δt, which will, therefore, be equal to the two way traveltime of the wave inside each layer. The true velocity of the EM waves within each layer is finite; in fact it is obvious that the EM waves will spend more time inside conducting media, than inside resistors, since they diffuse at a rate $1/\sqrt{\sigma}$. Therefore, there are no guarantees that each layer of thickness h_j , with $h_j \neq h_{j+1}$ in general, can be traversed at equal times. At least this would be physically unreasonable, yet this is what we observe. The paradox arises from the equal traveltime requirement, *imposed* by our sampling operations and is exactly what makes life 'not so rosy'. This paradox demands more investigation, because the frequency transform $Y(z)$ of the impulse response and the conventional definition of the admittance function as $Y(\omega) = H(\omega)/E(\omega)$ *are one and the same*, when derived from sampled data.

I propose that by sampling the data at steady rate, we constrain the system to respond within Δt time intervals. The system responds by:

(a) Distributing the energy loss (delay) equally over all time lags $j\Delta t$, consistently with its MD nature, and

(b) By allowing us to resolve the cause of the energy loss (i.e. the conductivity profile) with the constraint (uncertainty)

$$\sigma_j \, h_j^2 = \nu^2 ,$$

where ν is dependent on Δt. In more plain words the Earth system allows us to observe only 'pseudolayers', traversed by 'pseudowaves' travelling at a constant velocity, subject to the uncertainty above. Physically this can be explained by a forced linear superposition of waves coming from similar depth ranges. We constrain ourselves to observing the EM field at regular time intervals, thereby quantizing the system. We should appreciate that every time our data sampler reads an

EM field value, it picks up a field 'quantum' which includes contributions from several nearby secondary sources. By conjecture, our attempt to recover the conductivity profile will be subject to the same uncertainty, in either the frequency or the time domain. This defines to some extent another side of the non-uniqueness of the inverse MT problem, since, for a given sampling rate, there will be infinite combinations of conductivity and thickness products that produce the same real constant.

It is important to note that this uncertainty does not imply our inability to obtain a valid Earth response, and recover a physical geoelectric structure! Consider that there is no theoretical constraint or argument to necessitate that $u_j \neq u_{j+1}$ ($a_j(\lambda) \neq a_{j+1}(\lambda)$), or that the reflection coefficients do not vanish. The only requirement is that $a_j(\lambda) \geq 0$, for physical realizability. Therefore, the pseudolayers we observe may simply be subdivisions of a thicker conducting slab. These subdivisions are forced to produce (pseudo)responses from different depths, subject to the traveltime constraint. The reflection coefficient from these subdivisions may be vanishingly small, if they separate layers of equal electric parameters, and therefore, there will be no observable movement in the response curve.

In conclusion, it would appear that by considering phase field propagation and the observational uncertainties, we can argue that the 1D Earth can be approximated as a Goupillaud type medium.

The concept of equal traveltime (equal attenuation) has been discussed, in MT literature, only by Kunetz (1972) and Parker (1980). The next few paragraphs will be devoted to these important contributions. Kunetz, introduced his time domain interpretive analysis, based on the assumption of equal traveltime. His proposition is that the layered

Earth medium can be subdivided into a series of 'elementary layers' with constant ν^2 value; when adjacent elementary layers have the same conductivity, the reflection coefficients can be equal to zero, as discussed above. The Kunetz elementary layers and the observed 'pseudolayers', can coincide if an appropriate sampling interval Δt is chosen. This educated assumption can lead to a satisfactory resolution of the layered Earth structure by implementing Kunetz's procedures. The author is aware of the problem of time vs model resolution, and suggests that high sampling rates must be used when the Earth medium becomes more complicated (number of layers increases). Ideally, the sampling rate will be $\Delta t \rightarrow 0$, therefore the elementary layers will be infinitesimally thin, and the imposed delay will also be infinitesimal. This would give the upgoing (response) waves enough time to traverse the overlying structure and arrive at the surface without superposition and interference with other waves. For realistic and technologically affordable sampling intervals, the concept of 'high sampling rates' has yet to be specified. Kunetz's approach, as reappraised by Janoth (1984) produces very interesting results when applied to synthetic data, but it has never been tested in practice. I believe that it *demands* more consideration and development, as an alternative to conventional interpretation procedures.

Parker's proposition for a solution to the MT inverse problem is readily available and applicable. Parker bases his approach on Weidelt's (1972) analysis of $C(\omega)$. However, in contrast to Weidelt's highly theoretical approach (i.e. continuous conductivity distributions and smooth continuous response functions, that include the asymptotic EM field behaviour at zero and infinity), Parker assumes the pragmatic view, that we are faced with discrete, inconsistent data, extending to finite

bandwidths. Thus, he presents an analysis that is highly relevant to Weidelt's, but can also reduce to an inversion procedure applicable to real life data. Essentially, he constructs three types of solution spaces, which he names D^+, with discontinuous (δ-function) conductivity distributions, H^+ (a Goupillaud type medium), and C^{2+}, with continous conductivity distributions. The extensive and non-trivial nature of Parker's mathematics hinder a brief description, therefore, I will assume that there exists a basic degree of familiarity with his theory. The H^+ class of solutions is based on the argument that when the conductivity profile is an ordinary function, the Earth is electrically analogous to a cascade configuration (ladder) of uniformly distributed two-port RC networks (e.g. Ghausi and Kelly, 1968). The boundary conditions at the bottom of the system are approximated with a perfect conductor, therefore the analogue becomes a short-circuited RC network. Is this assertion reasonable? For the one dimensional case, the differential equation governing the behaviour of a two-port RC network (a coaxial transmission line) is

$$\partial V/\partial x - i\omega RC\, V = 0$$

where V is the harmonically varying potential, R is the resistance and C is the capacitance. Because we assume that $E \simeq V$, the physical analogy is immediately obvious. The transmission line analogue has been used extensively by several investigators, (e.g. Madden and Swift 1969; for a review see Ward et al, 1973, Porstendorfer, 1975). The principal characteristic of a distributed RC network is that the electrical length of each element of the system $(R\, C\, x^2) = a$, a constant. We see therefore that the concept of the transmission line has been used before in EM data analysis, and that the equation governing the EM field behaviour in the Earth and the transmission line are the same. Thus,

Parker's assertion is both reasonable and physically valid. The H^+ inverses reduce to an exercise of cascade synthesis of uniformly distributed RC networks, according to the procedures described by Ghausi and Kelly (1968), and subject to the physical constraints of the MT problem. On the other hand, I have argued that our data sampling operations impose the equal traveltime (equal attenuation) constraint, and therefore, uniform distribution properties on the estimated Earth response function. It appears that the observed response function and the H^+ inverses are subject to the same conditions and closely related; the response function is given with an uncertainty, and the H^+ inversion procedure tries to locate the optimum value of this uncertainty and the most probable conductivity profiles that obey it.

For real (digital band-limited) data however, the response function is calculated subject to a piecewise application of the data sampling uncertainty, per frequency band (or frequency decade). The complete frequency domain response curve is reconstructed as a piecewise colation of band limited response functions. Thus, in practical situations the uncertainty is not uniform over the whole frequency domain response. H^+ inversion assumes that it is. Therefore, the theoretical equivalence does not imply that the H+ inversion procedure is necessarily our best bet for reliable Earth parameter estimation. I think however, that it is very useful that we appreciate the existence of such an inherent relationship, which, in a sense, comprises a reappraisal of the much neglected H^+ model construction tool.

Some final notes apply with respect to the properties of the partial fraction expansion (6), and the piecewise discontinuous function u_j. Two results established so far, are relevant, i.e.

(1) the zeros of Z(z) correspond to the points of discontinuity of u_j

(2) the zeros of Z(z) and consequently, points of discontinuity of u_j correspond to the points of discontinuity of the conductivity profile. Now, consider that from standard spectrum theory we know that every spectrum $X(\omega)$ possesses a spectral distribution function $S(\omega) = S_1(\omega) + S_2(\omega)$, where $S_1(\omega)$ is a discontinuous deterministic non-decreasing function, and $S_2(\omega)$ possesses both discontinuous, and absolutely continuous non-decreasing components that may be deterministic or not. The deterministic component in the spectral function of $Z(\omega)$ is the finite order Earth stucture; the absolutely continuous component can be identified with noise. We have also seen that the *observable* Earth response appears to be given in terms of a discontinuous spectral function only. These facts are important in understanding the practical nature of the MT problem, because they allow some sweeping conclusions to be drawn about the form of the *recoverable* conductivity profile. The discontinuous nature of u_j can be attributed to the existence of an *equivalent* discontinuous conductivity profile, in a space that comprises a stack of thin conductive sheets, separated by insulating layers. The thin sheet model can be constructed by shrinking the conductivity of each layer (or pseudolayer) into infinitesimally thin sheets of finite conductance, at the layer boundaries. The response function is generated at the thin sheet, and energy dissipation contained there. The new profile will comprise N finite jumps, as

$$\sigma(d) = \sum_{j=1}^{N} \bar{\sigma}_j \, \delta(d-d_j),$$

where $\bar{\sigma}_j$ is the integrated conductivity (conductance) over the whole layered system, with the variable d taken positive upwards. The waves will be reflected and transmitted at depths $d = d_j$ and at no other

points; energy loss will, likewise be contained in the thin sheets, which can, therefore, be interpreted as *scattering (reflection) horizons*. This type of solution belongs to the class D^+ Parker (1980), Parker and Whaler (1981). In fact, Parker (1980) has shown that for layered Earth media, this type of conductivity distribution is fundamental to the MT inverse problem, and forms the basis for the construction of any other type of conductivity profile. Weidelt (1985) explores the possibility of constructing conductance bounds from magnetotelluric impedances, based on a very similar definition. A conductivity profile consisting of thin sheets is not a concept that will satisfy many investigators. It must be appreciated however, that the geophysicist's favourite uniform 'slab-stack' model, is inherently discontinuous, and changes with finite conductivity jumps. This is exactly what the thin sheet sequence implies; it is just the distribution of the conductivity that changes, i.e. it is integrated to conductance. Once we have accepted the existence of such a solution space, we can recover the thin sheet profile using Parker's machinery. According to the discussion above, this is the only available MT equivalent to a vertical seismic reflection profile, and does not constitute a 'weird' mathematical condition only. In our attempt to do so we may encounter various degrees of success. The numerical method is robust, but it tries to recover the 'reflectivity' profile from a frequency domain sounding curve; in addition to the known problems with the quality of estimation, this would be the wrong domain for such an exercise in the first place. In my experience however, the D^+ inverses can be an invaluable interpretational aid, because they will always tell us the whereabouts of the *conducting* parts of a layered Earth structure.

IV. The multidimensional case.

The 1D polynomial Earth response function discussed above, affords a generalization to higher dimensions, by extrapolation of the concepts and ideas discussed above. Then, it becomes a *matrix polynomial* (e.g. Robinson 1967). This discussion however, does not aim to utilize the concept in order to produce any theoretical development of the multidimensional MT problem; this would be a task beyond the scope of this work and is left to a later time. Here, I intend to provide a basic review of the properties of a minimum delay matrix polynomial, with the purpose of establishing some basic understanding for the behaviour and analytic properties of the Earth response function in higher than one dimension. I believe that this is a worthwhile task, since the existing literature on the subject is a null set! The only possible exception is a reference by Rokityansky (1982), to the work of Svetov et al (1978), which, however, does not appear to have been published in a language other that Russian. In the following, I shall assume the general case of a 3x3 matrix polynomial, providing a complete description of EM induction over a 3D conductivity distribution.

A matrix polynomial is generated by the linear relationship (convolution)

$$\mathbf{h}(t) = \mathbf{y}(t) * \mathbf{e}(t)$$

where $\mathbf{e}(t) = [e_x\ e_y\ e_z]$, $\mathbf{h}(t) = [h_x\ h_y\ h_z]$ and $\mathbf{y}(t)$ is, in general, a 3x3 matrix wavelet with the frequency transform

$$\mathbf{Y}(z) = \mathbf{y}_0 + \sum_{j=1}^{N} \mathbf{y}_j\, z^j$$

For a matrix polynomial there exist:

(a) The determinant of the z-transform, Det{Y(z)}.

(b) The adjoint matrix polynomial adj{Y(z)}.

(c) The inverse matrix polynomial $\mathbf{Z}(z) = \text{adj}\{\mathbf{Y}(z)\} / \text{Det}\{\mathbf{Y}(z)\}$.

The *determinantal equation* $\text{Det}\{\mathbf{Y}(z)\} = 0$, is a polynomial of degree at least $K \leq 3N$, with distinct simple roots $z_1, z_2, \ldots, z_K$. The roots of the determinantal equation are identified with the roots of the matrix polynomial $\mathbf{Y}(z)$. Obviously, for $z = z_j$, $\mathbf{Y}(z_j)$ becomes a singular matrix of rank 2. Then, $\text{adj}\{\mathbf{Y}(z_j)\} = \mathbf{c}_j \mathbf{r}_j$ is a 3x3 matrix of rank 1, with $\mathbf{c}_j$ and $\mathbf{r}_j$ being constant column and row vectors respectively. It follows directly from (c) that, to each root z_j of $\mathbf{Y}(z)$, one can assign a column vector $\mathbf{c}_j$ and a row vector $\mathbf{r}_j$, such that

$$\mathbf{Y}(z_j)\ \mathbf{c}_j\ \mathbf{r}_j = 0$$

Thus, we shall call the roots z_j the eigenvalues (eigenroots) of $\mathbf{Y}(z)$; $\mathbf{c}_j$ and $\mathbf{r}_j$ shall be the column and row eigenvectors of the matrix polynomial $\mathbf{Y}(z)$. It can be seen immediately that things are considerably more complicated than in the 1D case, since the roots of the polynomial are more numerous than its order.

The complication is quite severe when one considers the factorization of a matrix polynomial in terms of binomial factors. This is permissible (e.g. Robinson 1967, Claerbout 1976) and assumes the form

$$\mathbf{Y}(z) = \mathbf{b}_0\ (\mathbf{I} + z\mathbf{b}_1)\ (\mathbf{I} + z\mathbf{b}_2)\ldots(\mathbf{I} + z\mathbf{b}_N)$$

where $\mathbf{b}_n$ is a constant 3x3 matrix. The following results apply on the basis of such a factorization:

(1) If $\text{Det}\{\mathbf{Y}(z)\}$ is a minimum delay polynomial, then, $\mathbf{Y}(z)$ is minimum delay and vice versa. Then, each factor $(\mathbf{b}_n, \mathbf{I})$ is also minimum delay. Notably, there exist $(3N!)/(N!)$ ways to factor a 3x3xN multivariate spectrum (matrix polynomial), but the minimum delay factorization is unique.

(2) If $\mathbf{Y}(z)$ is minimum delay, then $\mathbf{Z}(z)$ is also minimum delay, and vice versa. Derivations of these propositions can be found in Robinson

(1967).

Finally, the definition of the inverse polynomial (c) gives

$$\mathbf{Z}(z) = \text{adj}\{\mathbf{Y}(z)\} \; / \; y_K \, (z-z_1)(z-z_2)\ldots(z-z_K),$$

so that, in analogy with the 1D case there exists a partial fraction expansion of the form

$$\mathbf{Z}(z) = \sum_{j=1}^{K} \frac{\mathbf{u}_j}{(z - z_j)}$$

where

$$\mathbf{u}_j = \text{adj}\{\mathbf{Y}(z_j)\} \; / \; y_K \, (z - z_1)(z - z_2)\ldots(z - z_K) = \text{adj}\{\mathbf{Y}(z_j)\} \; \mathbf{a}_j .$$

Thus, the equivalent partial fraction expansion for the minimum delay matrix polynomial $\mathbf{Z}(z)$ becomes

$$\mathbf{Z}(z) = \sum_{j=1}^{K} \frac{\mathbf{a}_j \mathbf{c}_j \mathbf{r}_j}{(z - z_j)}$$

From the above, we can readily establish the analytic properties of the multivariate impedance function $\mathbf{Z}(z)$. The minimum delay requirement imposes the property $|z_j| > 1$. Thus, the roots of the polynomial are located outside the unit circle. The exterior of the unit circle is mapped onto the upper-half ω-plane. Therefore, if by a limiting operation we convert the z-transform to Fourier transform, we establish the fact that all the zeros of the matrix $\mathbf{Z}(\omega)$ exist in the upper half ω-plane. Also, the requirement that $\mathbf{z}(t)$ be a positive real matrix wavelet with a positive definite cosine transform and a positive definite determinant, necessitates that the zeros of $\mathbf{Z}(\omega)$ be located on the negative imaginary axis. The dispersion relations follow from the property of the infinite order minimum delay determinantal polynomial, in the usual manner (e.g. Fischer and Schnegg, 1982). The properties of the multidimensional transfer function $\mathbf{Z}(\omega)$ can be understood from filter theory, without the need of analysing a multidimensional Sturm-Liouville problem.

It can easily be seen that

$$\mathbf{Y}(z) = (\mathbf{E}(z)\mathbf{H}(z)^{*T})(\mathbf{H}(z)\mathbf{H}(z)^{*T})^{-1}$$

which is equivalent to a 'matrix rational function'. From this expression, one can construct a multidimentional extension of the scattering formalism (10c). This involves the extension of the reflection and transmission coefficients to a 3x3 tensor form, so that scattering in three dimensions can be described. There exist three eigenroots per matrix coefficient $\mathbf{y}_j$, that provide information about the magnitude and phase transitions of the EM waves; spatial information is provided by the eigenvectors $\mathbf{c}_j$ and $\mathbf{r}_j$ of the matrix coefficient $\mathbf{y}_j$. The formalism possesses some very non-trivial features, and is currently under investigation. It is hoped that some progress can be reported in the future. Detailed information about inverse scattering theory in multidimensions can be found in Newton (1984) and references therein. The reader may also be interested to know that one interpretational method that explicitly involves wave propagation and inverse scattering in two dimensions, has already been developed and implemented by Lee et al (1987). It is called 'Phase field Imaging' and is the EM analogue of seismic migration.

CHAPTER 2

PART 2:

A SPECTRAL STUDY OF MAGNETOTELLURIC DATA AND NOISE INTERACTIONS

V. Introduction

Conventional magnetotelluric data analysis requires the solution to the frequency domain equations:

$$E_x(\omega) = Z_{xx}(\omega)\, H_x(\omega) + Z_{xy}(\omega)\, H_y(\omega)$$

$$E_y(\omega) = Z_{yx}(\omega)\, H_x(\omega) + Z_{yy}(\omega)\, H_y(\omega)$$

or, in matrix form, $\mathbf{E} = \mathbf{Z}\,\mathbf{H}$, where $\mathbf{Z}$ is termed the impedance tensor or response function, and conveys information about the geoelectric structure in the vicinity of the recording station. The field components $\mathbf{E} = [E_x\ E_y]$ and $\mathbf{H} = [H_x\ H_y]$ are recorded in the time domain and form a mutually orthogonal set. The two pairs of impedance elements are usually estimated by least-square solutions which minimize noise on a particular data channel (Sims et al, 1971). The quality of the least square solutions for each of the two pairs of impedance elements is then obtained from the multiple (predicted) coherence between the measured electric field component and that predicted by the least-square solutions.

The auto- and cross-spectral estimates between the field components are conventionally computed using the Fast Fourier Transform (FFT) technique. Given the inherent variance of the resulting raw spectral density function, smoothing over a frequency interval, often arbitrarily chosen, is usually prescribed, in order to stabilize the spectral density estimates. The frequency resolution of the resulting impedance

tensor estimates is thus drastically reduced. High frequency resolution will not, theoretically, offer additional information, because the impedance tensor (i.e. the Earth response), is a smooth function of frequency. However, no studies or examples of high resolution impedance estimation and its associated properties have appeared in the literature. In addition, a highly resolved Earth response will clearly display the effects of spectral components that are extraneous to the passive induction process. This may ultimately assist in their identification and removal.

Noise problems are accentuated for data obtained in the audio and sub-audio frequencies. Here we are not only faced with the natural non-stationarity of the EM field, but we encounter anthropogenic fields as well. The power distribution grid and its users contaminate the data with intense extraneous noise sources. The noise structure of a well balanced grid supply will be time-stable and comprise a fundamental (50Hz or 60Hz) and associated line harmonics. Variable loads on an unbalanced grid supply provide time-variable line sources and wide band contamination. Powerful irregular transients, also exist to provide additional narrow and wide band contamination.

In this study, I investigate some high resolution properties of noise contaminated EM data. Attention is focused on data from the lower audio and sub-audio frequencies, the so called AudioMagnetoTelluric (AMT) data, as they are both more affected, and of more interest for upper crustal geoelectric studies. To obtain highly resolved and stable impedance estimates I have adopted the Maximum Entropy (MAXENT) spectral analysis method. A detailed account on its implementation and rationale is given in Appendix 2.A.

In order to understand the behaviour of an Earth response function contaminated by noise, I shall make use of the definition of the response function as a finite degree polynomial, provided in PART 1. AMT data in the frequency range 0.1 to 100 Hz obtained at a number of sites in western Turkey are used to provide both conventional and high resolution estimates of the impedance tensor. Although the two sets of results can be considered equivalent for certain (low-noise) data types, the high resolution estimates display frequency-local structure for high-noise data. The results obtained are used to to identify a number of 'inconsistent' features that arise due to the contaminating influence of noise. Noise sources, structured in frequency, appear to influence several portions of the complete observational bandwidth. High resolution spectral density functions of the data are also used to assist in our understanding of noise structure which appears to be generated by both harmonic and subharmonic processes.

The piecewise distortion of the response function we observe in many of the high resolution estimates is due to the convolution of the time Earth response with inconsistent noise structure in several portions of the spectrum. All such discontinuous portions imply an unphysical Earth response and identify inconsistent sections of the spectrum. The recovery of a consistent high resolution response from such data requires careful consideration. Any attempt to remove noise contamination in the time domain is only effective if the noise structure is known *a priori*. For measured data, the noise structure is complex and it can only be understood *a posteriori*, using the high resolution procedures described in Appendix 2.A. Any attempt to remove piecewise distortion in the frequency domain involves a smoothing of the response function. The smoothing function must be consistent with known

properties of the response function and it must not impose any other structure on it with no a priori information. A smoothing operation, involving convolution in the frequency domain, is suggested and applied to the results. The technique exploits the minimum delay properties that must be exhibited by a valid Earth impulse response function in the time domain.

VI. Noise processes and the polynomial Earth response.

Having acknowledged the nature of the finite impedance polynomial and its strict dependence on the geoelectric structure, (PART 1), it is straightforward to understand the effects of non-physical frequency variations introduced by noise, because the observed Z(w) departs from its natural (low) degree as it is forced to adapt to the variable properties of the frequency domain. Thus *additional* poles and zeros are introduced. To investigate how this happens, let $N_E(\omega)$ and $N_H(\omega)$ be noise processes affecting the electric and magnetic field components respectively. Noise processes can be classified as strictly (second order) or weakly (first order) stationary. The first kind refers to additive white (random) noise. The second kind may incorporate all the natural and anthropogenic processes that are not due to EM induction, such as ground roll affecting the stability of the magnetic sensors, and the power distribution grid and its users. In all such cases the noise spectrum is expressible in terms of a polynomial (e.g. Robinson, 1980 p.187), not necessarily minimum phase or narrow-band. I will consider $N_H(\omega)$ and $N_E(\omega)$ to be a mixture of both kinds of noise processes. Then the operation

$$Z(\omega) = \frac{E(\omega) + N_E(\omega)}{H(\omega) + N_H(\omega)}$$

defines a *convolution* in the time domain, of $z(t)$ with $n_h(t)$ so that

$$e(t) + n_e(t) = z(t) * h(t) + z(t) * n_h(t) \qquad (13)$$

The contaminating agent emerges as a polynomial of degree *at least* equal to that of $Z(\omega)$ i.e. the polynomial to which it has been added. If the noise sources define a polynomial of degree greater than that of $Z(\omega)$, (13) defines how additional zeros are introduced to the observed response. The application of the parallel filter rule (e.g. Claerbout 1976, Robinson 1980) tells us what the relative effects of $N(\omega)$ are on $Z(\omega)$. The rule states that the frequency transform of the two processes has the same number of interior (to the unit circle) zeros, (or equivalently the same number of zeros in the upper half ω-plane) as the transform of the dominant process. In particular, if the dominant process is minimum delay, then the parallel combination of the processes will be minimum delay; if it is mixed delay, the combination will have a number of zeros in the upper half ω-plane and instability will occur. In other words, "You can add garbage to a minimum delay wavelet if you do not add too much" (Claerbout, 1976). In normal circumstances, not too much is added, and a minimum phase $Z(\omega)$ is recoverable. In the extreme case however, $Z(\omega)$ will be contaminated by dominant mixed phase processes introducing instability and frequency-local variations, as well as white noise (i.e. a lot of garbage).

A conventional way of reducing the variance in standard frequency domain processing is to smooth the observed auto- and cross-spectral estimates of E and H by averaging over a usually arbitrary frequency interval. Such a technique of stabilizing the spectrum however is only adequate if the contaminating agent is a second order stationary process, the effects of which are expected to cancel upon averaging. This cannot be claimed if the noise terms are defined as structured (polynomial)

processes; then, averaging is inadequate. The properties of the noise spectral components will be included in the 'smoothed' spectrum, and even worse, if they are narrow band and sufficiently powerful, they will disturb neighbouring frequencies. When dealing with such problems, our ability to identify the extent to which our data have been affected becomes important and brings forward the question of frequency resolution in obtaining an Earth response. The problem of eliminating the noise structure is then left to a second stage procedure. This study will be concerned with our understanding of an Earth response and its recovery from data contaminated by a lot of garbage.

VI.1 The AMT data.

In this study we are primarily interested in the properties of data in the lower audio- and high sub-audio frequencies (0.1Hz <f< 100Hz). The data were acquired with a minicomputer based AMT field system covering the bandwidth 0.01 to 100Hz. Band-pass and notch filters were applied to condition the analogue signals prior to 12 bit digitization. The conditioning scheme provided four decades of data with sampling rates of 400, 40, 4 and 0.4Hz across a window of 300 data points per decade. The data are resampled at 200, 20, 2 and 0.2Hz across a window of 150 data points, so that the four decades cover the bandwidths 100-10Hz, 10-1Hz, 1-0.1Hz and 0.1 to 0.01Hz respectively. The time interval for the collection of each data window is thus 0.75, 7.5, 75, and 750 seconds respectively. Data recording was discontinuous due to in-field processing and storage on a slow digital cartridge. Collection took place in western Anatolia (Turkey), geographic coordinates 40.5°N, 30°E.

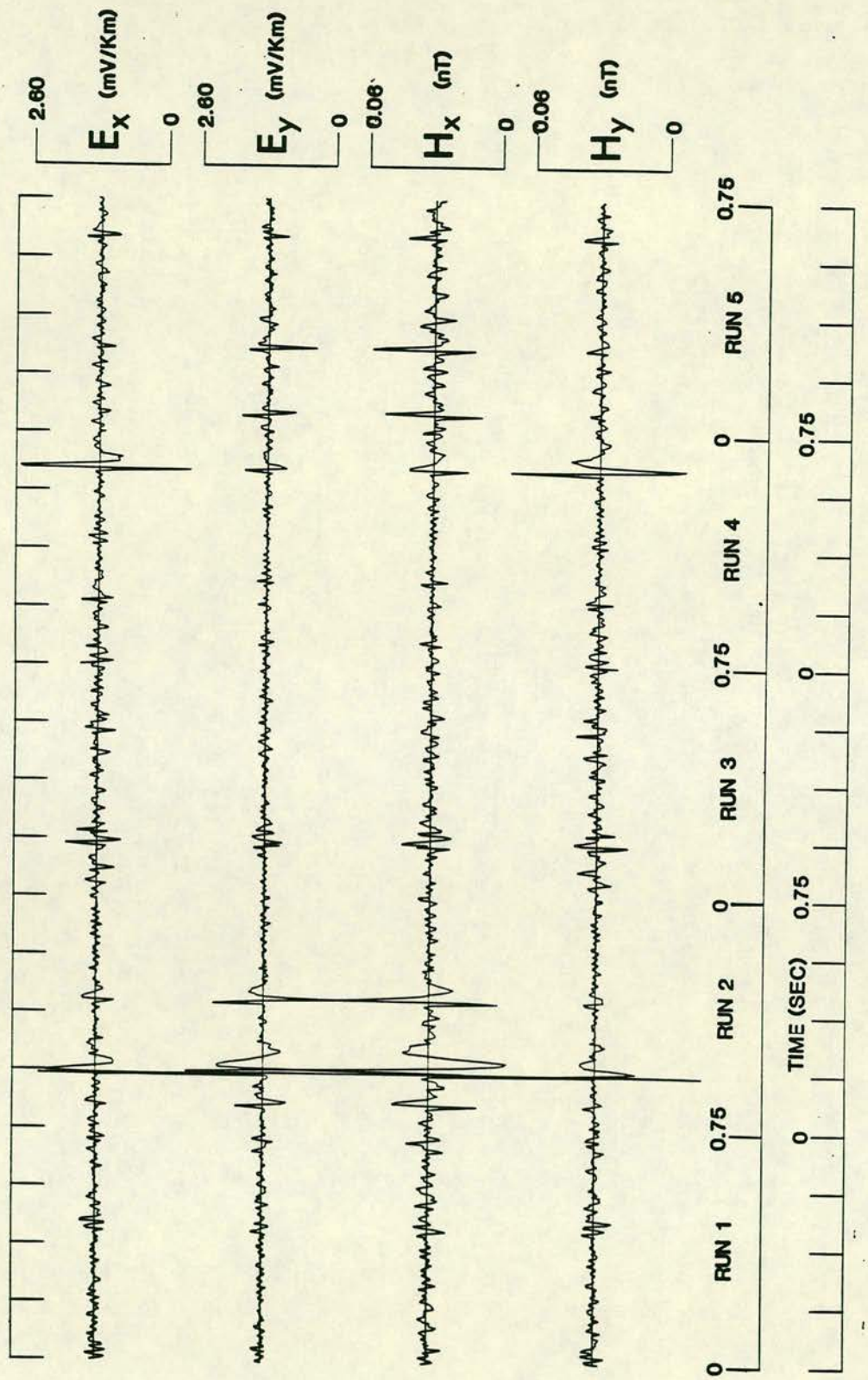


FIGURE 1. Four channels of decade 1 data. Sampling rate is 200Hz. Five successive data windows of 0.75 second duration. E_x and E_y are the N-S and E-W orthogonal components of the induced electric field respectively; H_x and H_y are the corresponding magnetic N-S and E-W components.

Figure 1 displays example Decade 1 data (100-10Hz) for five successive data windows (RUN numbers), in two orthogonal telluric and two orthogonal magnetic channels. The 100-10Hz bandwidth is dominated by the Schumann resonances waveforms; in general they consist of transient events (sferics) representing the response of the Earth-ionosphere waveguide to extra large lightning discharges, superimposed on a background due to the continuous global sum of lightning activity (e.g. Bliokh et al, 1980). It is evident that this data can be approximated by a sequence of locally stationary processes at best.

VII. High resolution Earth response functions and their properties.

In the following I shall give examples of the properties of high resolution Earth response functions and compare the results with those based on the conventional FFT approach, for data recorded at different dates and at different locations.

The spectral parameters used in the analysis here are, (a) For MAXENT: data lengths of N=150 fitted by AR models of order M=10, unless otherwise stated, (b) For FFT: N=150, cosine-tapered and extended to N=256 by symmetrically appending zeros; variance reduction by averaging over 12 adjacent frequencies provides 9 response estimates per decade associated with 16 real degrees of freedom. For the comparisons, data processing (other than spectral analysis) was *identical*. The impedance tensor elements were calculated by forming the weighted means of the individual element populations that passed a predicted coherence threshold of 0.80. Weights were the random errors associated with each element. Upward- and downward-biased elements were treated separately

and subsequently combined to produce the 'unbiased' result displayed. I present the modulus of the unrotated off-diagonal impedances (logarithmic scale) and their associated phase response as a function of frequency (logarithmic scale). Errors refer to one standard deviation throughout.

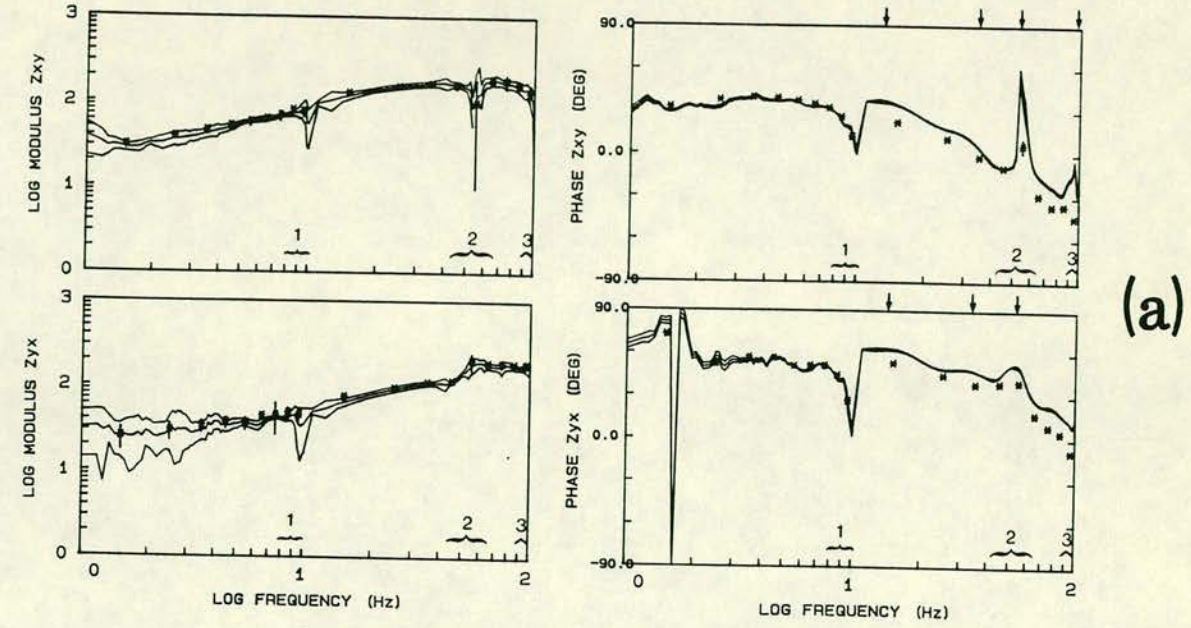
VII.1 Example Site 1.

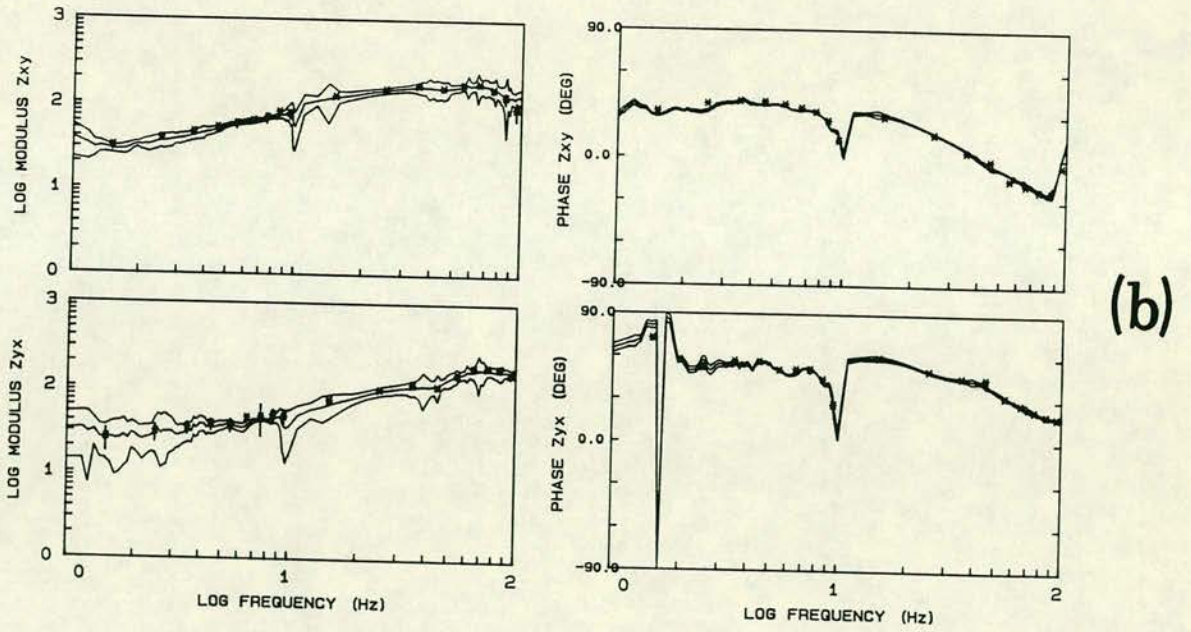
Two decades of the off-diagonal impedance function from site 1 are presented in Figure 2. The main inconsistent features observed are indicated by numbers 1, 2 and 3. **Feature 1** broadly covers the 8-10Hz range. It is less evident in the modulus of the impedance function and more impressive in the phase response that appears to be sloping downwards; the unconformity between the two decades is apparent in the divergence of the phase response. This is a combined effect caused by instrumental filters, and, mainly, by (the principal) aliases of the power grid fundamental. The phenomenon requires special attention and will be discussed later in more detail. **Feature 2** is the effect of a strong residual noise component due to the power distribution grid that leaked through the sidelobe structure of the 49Hz instrumental notch filter because of its unstable harmonic content. Thus, it can be seen to affect the bandwidth between 43 and 62Hz, and, is apparently more extreme in the Z_{yx} component, despite the impressive spectral lines observed in Z_{xy}. Feature 2 is a case of noise differentially affecting the field components. **Feature 3** is a scaled down version of Feature 2, which is believed to be due to the first harmonic of the unstable mains fundamental. It is centered at 97.9Hz, affecting the 92-100Hz bandwidth; it possesses similar time dependent characteristics. The

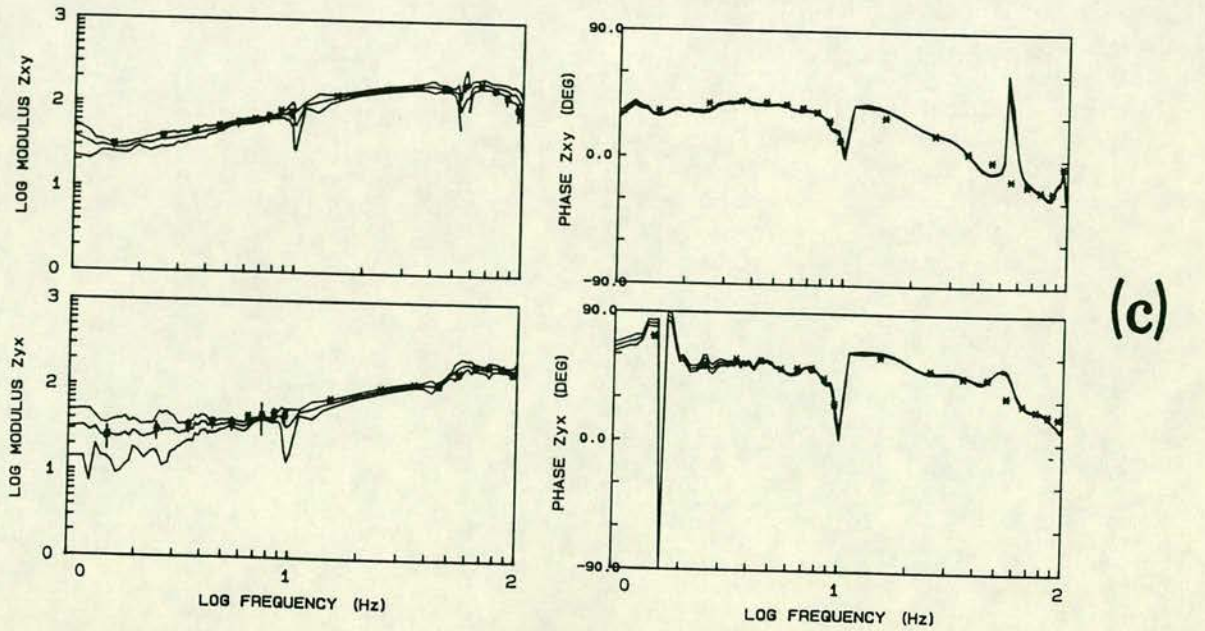
qualitative and quantitative difference on the two elements is obvious.

There is another interesting effect that can be resolved with such a data adaptive technique. Consider Figure 2, in which the solutions obtained with both MAXENT and FFT techniques are compared. In Figure 2a, the decade 1 impedance has been evaluated from data affected by the mains harmonics. In Figure 2b the decade 1 impedance has been evaluated from 'cleaned' data with the mains harmonics and their sidelobe structure removed through time domain inverse filtering; details of the procedures are described in Chapter 3. In Figure 2c, solutions obtained with MAXENT for *noisy* data are compared with those obtained with FFT for *clean* data. The decade 2 solutions are the same for both techniques. It is apparent that the decade 2 solutions obtained with both techniques are in good agreement and display a structured spectrum for both the amplitude and phase response, indicating frequency local interference. Note the behaviour of the 1-3Hz bandwidth in particular. As can be seen in Figure 2a the decade 1 *amplitude* response functions evaluated by either spectral method agree, apart from the 10-20Hz bandwidth. This is not the case for the *phase* response. We observe that the FFT phase appears to be downward shifted with respect to the MAXENT phase. From Figure 2b we observe that this effect no longer occurs for *'clean'* data, and the phase responses are in very good agreement. Furthermore, from Figure 2c, the *clean* FFT response and the *noisy* MAXENT response match fairly well. It would appear that the phase response obtained with FFT for the *noisy* data is an incorrect function, while the MAXENT phase is more consistent. The downward pointing arrows in the phase diagram of Figure 2a indicate the average positions of the strong residual (53.4Hz) mains component and what appears to be the effect of its associated sideband and/or subharmonic structure. They are seen as smooth

FIGURE 2. Demonstration of the performance of MAXENT spectral analysis in AMT impedance estimation - comparison with FFT based analysis, for site 1 data. Bandwidth shown extends from 1-100Hz. Comparisons of impedances computed: **(a)** from the original field data contaminated by noise due to the power distribution grid; **(b)** from the same data but with noise affecting decade 1 removed; **(c)** from contaminated data (MAXENT) and 'clean' decade 1 data (FFT). The continuous curves correspond to the high resolution (MAXENT) response functions; the stars to the discrete (FFT) response estimates.


LOG MODULUS Zxy
LOG MODULUS Zyx
PHASE Zxy (DEG)
PHASE Zyx (DEG)
LOG FREQUENCY (Hz)
1
2
ω
(a)


LOG MODULUS Zxy
LOG MODULUS Zyx
PHASE Zxy (DEG)
PHASE Zyx (DEG)
LOG FREQUENCY (Hz)
(b)


LOG MODULUS Zxy
LOG MODULUS Zyx
PHASE Zxy (DEG)
PHASE Zyx (DEG)
LOG FREQUENCY (Hz)
(c)

undulations superimposed on the 'true' phase movement in the vicinity of 12, 30, 50 and 98Hz. It must be appreciated that resolution down to the level of the noise spectral lines is not possible because of their low power level and the dispersive filtering effect of the Earth. What we are actually observing is their combined interference effect. It could be that spectral averaging over a bandwidth containing a rapidly moving phase, and structured noise components, introduces harmonic distortion and phase shifts. When we compare the response obtained by the two spectral methods for the cleaned data set however (Figure 2b) we see that they are almost consistent. For noise-free, consistent data, the determination of the EM response functions should be independent of the spectral analysis technique and this is what we observe.

VII.2 Example Site 2.

The data obtained at site 2 are affected by a moderate mains harmonic (50Hz) and its associated higher and lower order structure, with a stronger component in the E-W direction. The quality of the data was good with more than 70% of the recorded data windows possessing predicted coherences >0.80. Two decades of the off-diagonal impedance function from site 2 are presented in Figure 3. The amplitude response for this site is a smooth, well-behaved function of frequency, but again, the most interesting features appear in the highly structured (ripple-like) phase response, particularly in the first decade (Figure 3a). Here as well, the crests of the ripples are believed to correspond to the effects of the mains associated harmonic structure (near 10, 30, 50 and 100Hz), subject to the same resolution constraints as per site 1. An additional feature is present at about 70Hz and is of unknown origin.

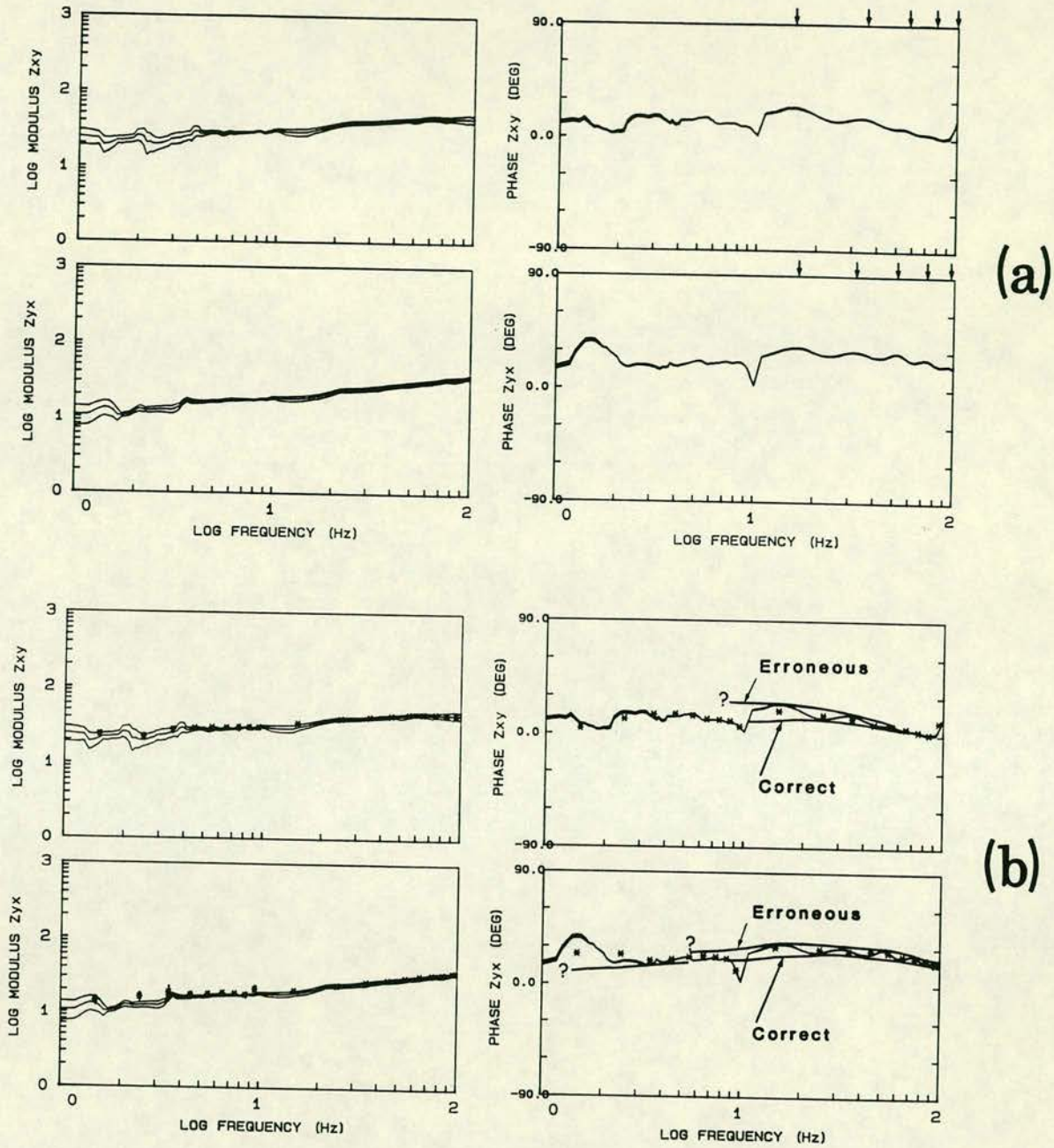


FIGURE 3. As per Figure 2 but for site 2. **(a)** High resolution impedance functions from data contaminated by weak mains noise sources. **(b)** Comparison between MAXENT and FFT response functions.

The troughs again correspond to the relatively unaffected frequency components. However, the fundamental (50Hz) harmonic appears very weak, in contrast to that at site 1. The noise effects again are more intense in the Z_{yx} tensor element and become more pronounced because of the relatively flat 'true' phase response upon which they are superimposed. Decade 2 displays the same 8-10Hz distortion as per site 1. Structure is also observed in the decade 2 Z_{yx} tensor element with properties similar to those of site 1.

Figure 3b displays the correspondence between FFT and MAXENT derived impedance functions. A good agreement between them is immediately apparent in both amplitude and phase. The flat and smooth nature of the latter does not present us with shifting problems as for site 1. The FFT phase estimates appear to be consistent averages of a continuous (MAXENT) response curve. This however means that they contain erroneous as well as valid contributions, and as a result, they are *not* correct estimates of the true phase. The frequency components near the troughs of the ripple structure apparently provide more reliable estimates of the true phase response. In fact two asymptotic bounds can be drawn for the MAXENT phase data, as shown in Figure 3b. These correspond to what I call the 'correct' and 'erroneous' response, and can be compared with the FFT results. The differences are not trivial; if we are in a position to obtain error bounds accurate to within 5%, any modelling of the response obtained may produce misleading results if incorrect estimates are used.

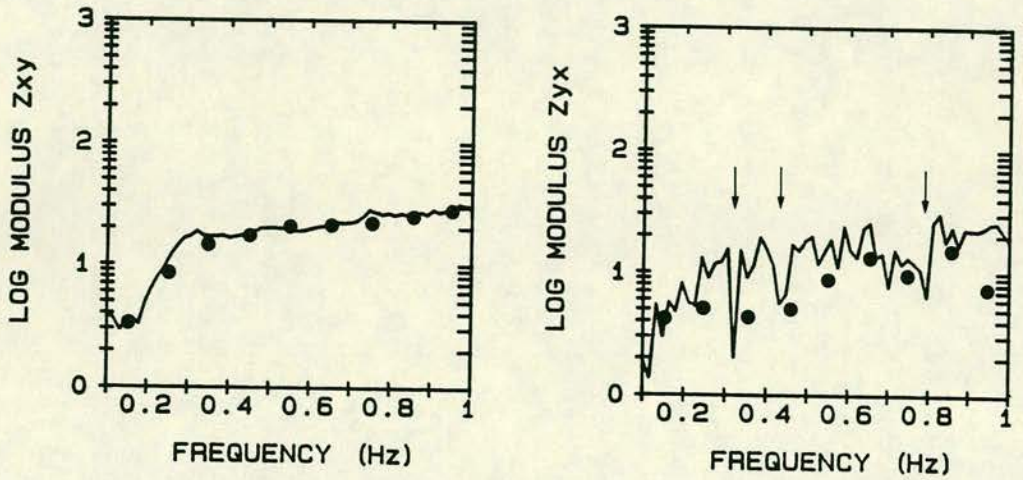
VII.3 Example Sites 3 and 4.

The third example aims to emphasize some difficulties that can arise

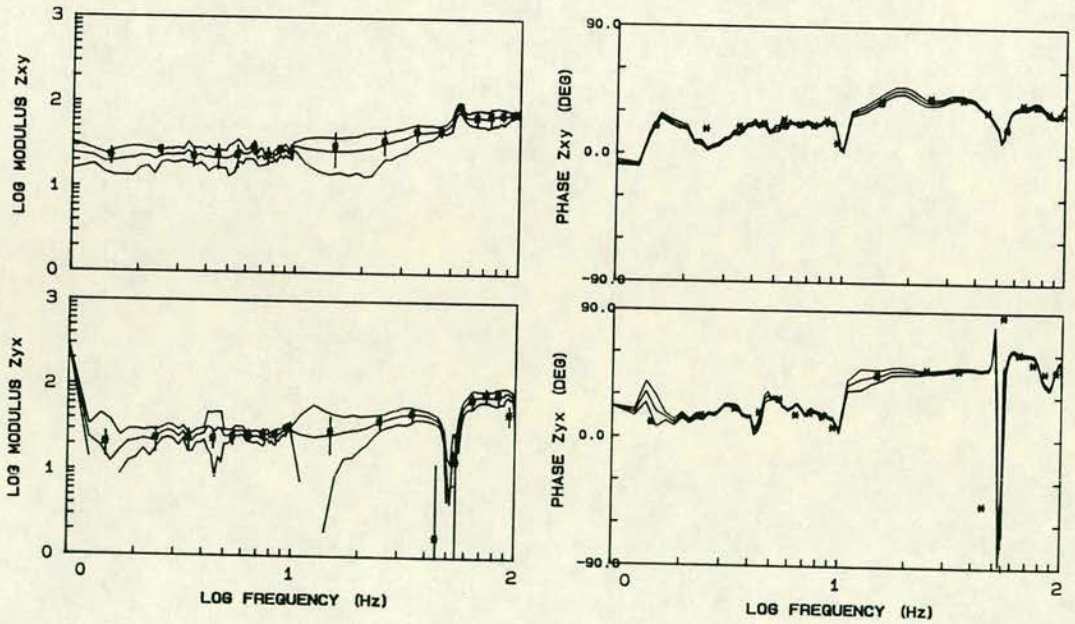
when the conventional method smoothing the auto- and cross-spectral estimates, by band averaging, is applied in cases of severe frequency local noise contamination. The example selected for demonstration, belongs to decade 3 data, and was recorded in a very high noise environment, with very directional (E-W) characteristics. The character of the noise signals was dual; it comprised random, low amplitude spikes, and structured harmonic processes at the vicinity of 0.1Hz, 0.32Hz, 0.45Hz, and 0.7Hz-0.85Hz. Such noise structure is very similar to that observed for the decade 3 data of site 1, the received spectra of which are displayed and discussed in Figure 6. The modulus of the Z_{xy} and Z_{yx} elements of the impedance tensor are shown in Figure 4a; where overlayed plots of the high resolution (MAXENT) and the FFT results are presented. For the sake of clarity and simplicity, error estmates and phase responses are not shown. One can again observe the self-consistent nature of the MAXENT and FFT results, for the adequately estimated Z_{xy} element. However, things are very different for the Z_{yx} element. The location of the structured noise components is evident from the large amplitude spikes, and is pointed by the arrows. Next, observe that the averaged FFT spectra display considerable downward bias due to the consistent nature of the spectral contaminants; band averaging over closely spaced noise components allows the noise to predominate over the data. The MAXENT spectrum does not, of course, avoid these complications. It can be seen that the vicinity of the 0.1Hz band is totally corrupted for both Z_{xy} and Z_{yx} components. We make this inference because we consider the steep gradient of the impedance to be unphysical, over such a narrow bandwidth of 0.2Hz only. Except for this common feature, the rest of the Z_{xy} component is a smooth and stable function of frequency. In the Z_{yx} element, we observe

that several portions of the bandwidth are left relatively unaffected by the structured noise manifestations, and are, therefore, marginally recoverable. Its high resolution estimation allows for a 'frequency local recovery' of the adequate data properties. One other notable feature is observable in the 0.7Hz-0.85Hz of the Z_{yx} element. With the noise sources affecting a wider bandwidth, both the MAXENT and the FFT methods experience similar distortion, and return similar downward biased, self-consistent results. The instability observed in the remaining portions of the bandwidth is due to the random noise components.

Note that in the above discussion, I do not mean to suggest that the FFT spectral technique is an unwarranted procedure to apply when 'structured' noise spectra are present. I demonstrate these investigations by presenting a few rather special cases, as extremal data distortion is exactly the issue in this study. In all normal circumstances, and for most 'noisy' cases, the two techniques return consistent results. In order to reassert this point, at least for the harmonic noise problem, in Figure 4b I present overlayed plots of the impedance functions returned by both spectral techniques for a different site, Example Site 4. The site apparently suffers from more extreme noise conditions than the first two examples, but both techniques return self-consistent results. This indicates that it is the local combination of noise and data statistics that determines whether problems such as those described above will actually occur. This is somewhat unfortunate because no prior information about such interactions is possible and defines the degree of caution that must be exercised while analysing and interpreting data recorded in intense noise regimes.



(a)



(b)

FIGURE 4. **(a)** The modulus of the high resolution (continuous line) and the discrete (FFT) impedance function (solid circles), at a third example site. No error estimates are shown.

(b) The modulus and phase of the high resolution (continuous line) and the discrete (FFT) impedance function (stars) at a fourth example site.

VII.4 Correlations

In the above presentation I have demonstrated the existence of structure in the impedance spectra, and, interpreted it as the result of man made frequency-local noise interference. It has also been acknowledged that part of this interference is due to the mains harmonic and subharmonic structure. In this section I shall attempt to investigate the causes of these effects by comparing the 'received' characteristics of the data.

One basic question is how can subharmonic interference be produced, or, equivalently, what is the frequency divider(s) producing it? The explanations I can offer are limited to the level of my understanding of the processes involved in the complex pattern of noise generation and propagation. However, I am in a position to present evidence corroborating subharmonic generation due to an unstable power distribution grid. Figures 5a and 5b show decade 1 and 2 data obtained at a fifth location. The two components of the telluric field are presented. The data were recorded quite near to a power distribution substation. For these data, the instrumental (50Hz) notch was applied to the E_x component but *not* the E_y component. The enormous 2.3V/Km sinusoid observed along the E-W direction (E_y) is therefore the *actual* signal emitted by the grid. The N-S (E_x) component displays the effects of the application of the instrumental (50Hz) notch filter, intended to counter this particular noise problem. In addition, the noise along both directions appears to be *amplitude modulated*. Figure 5c displays the MAXENT spectra of an accumulation of 165 data windows for both decades, in the E_x and E_y components. The spectrum of the fundamental mains harmonic can be seen to exist in two distinct modes, at 48.8Hz and 51Hz (E_y component). These modes could possibly be interpreted as

eigenfrequencies of an unstable mains grid, resulting from short term variations of the fundamental. In turn, they can be used to explain most of the features observed in the decade 1 data, as they will provide a superposition of waves travelling with different relative frequencies. The asynchronous operation of motors can produce more sidebands and subharmonics of the mains emissions (e.g. Lokken and Shand, 1964) and contribute towards this effect. The net result will be the emergence of a wavepacket displaying amplitude modulation as observed in the data, and the generation of the 98Hz harmonic, as well as a series of subharmonics. The n=2 (25Hz) subharmonic is clearly observable in the E_x spectra, while both the n=2 and n=3 (12.5Hz) subharmonics are evident in the E_y component (note the scale size). It must also be appreciated that the power associated with this subharmonic structure will be variable with time, and associated with the propagation properties of the wavepacket. Therefore, their effect can be *underestimated* in such a time-averaged spectrum.

The emerging wavepacket appears to travel with a given group velocity on a carrier frequency with unstable characteristics. As a result, noise power could leak into lower frequency bands, in the form of a wavepacket envelope. Such an effect can be observed in the 4-8Hz frequency range of Figure 5c. Recall that the E_x telluric component was recorded with the instrumental notch operational for both decades. The reduction of the power level at about 50Hz, causes considerable reduction of leakage into decade 2; hence the qualitative difference with the E_y component, recorded with the notch 'out'. An approximately 2 order of magnitude difference in the power level is evident, while the shape of the spectrum approaches the theoretically expected one, that must display a minimum in this frequency range. The instability of the power

distribution grid appears to provide the frequency dividers required for subharmonic generation.

Removing the leaked power however, reveals a structured spectrum, stable throughout the decade, correlated in the two channels, and, unaccounted for by the theoretical predictions for the natural EM field. I believe that these are effects of cultural activities superimosed on the natural spectrum. I cannot be certain about their sources and propagation mechanisms, but in the light of the above discussion the effects of the power distribution grid to lower frequencies cannot be ruled out.

Consider now the spectral line observed in the decade 2 E_y telluric component, at 8.8Hz (indicated by the arrow). This feature cannot belong to the fundamental Schumann resonance mode, because (a) it is too powerful, and (b) its Q factor is too high (>20). The Q factors of the Schumann resonances reflect the very low transmission coefficient of the ionosphere; they should in general be < 8 (e.g. Bliokh et al, 1980). I believe that this spectral line is adequately interpreted as the principal (n=1) alias of the 48.8Hz mains mode. The folding of aliasing frequencies into lower frequency bands is given by the expression

$$f_a = f - 2mf_N , \qquad m=1,2,...$$

with f_N representing the Nyquist frequency. For decade 2 data f_N = 20Hz during data collection. The generation of the aliased component at 8.8Hz can thus be acknowledged. Again it is interesting to observe the effect of the notch filter on the aliased power level of the E_x telluric channel. The power of the 48.8 Hz peak is reduced by almost 4 orders of magnitude. Accordingly, the aliased power level is drastically reduced. However, enough of it remains to fold back into decade 2 and form a low amplitude aliased spectral component. This is important in that it

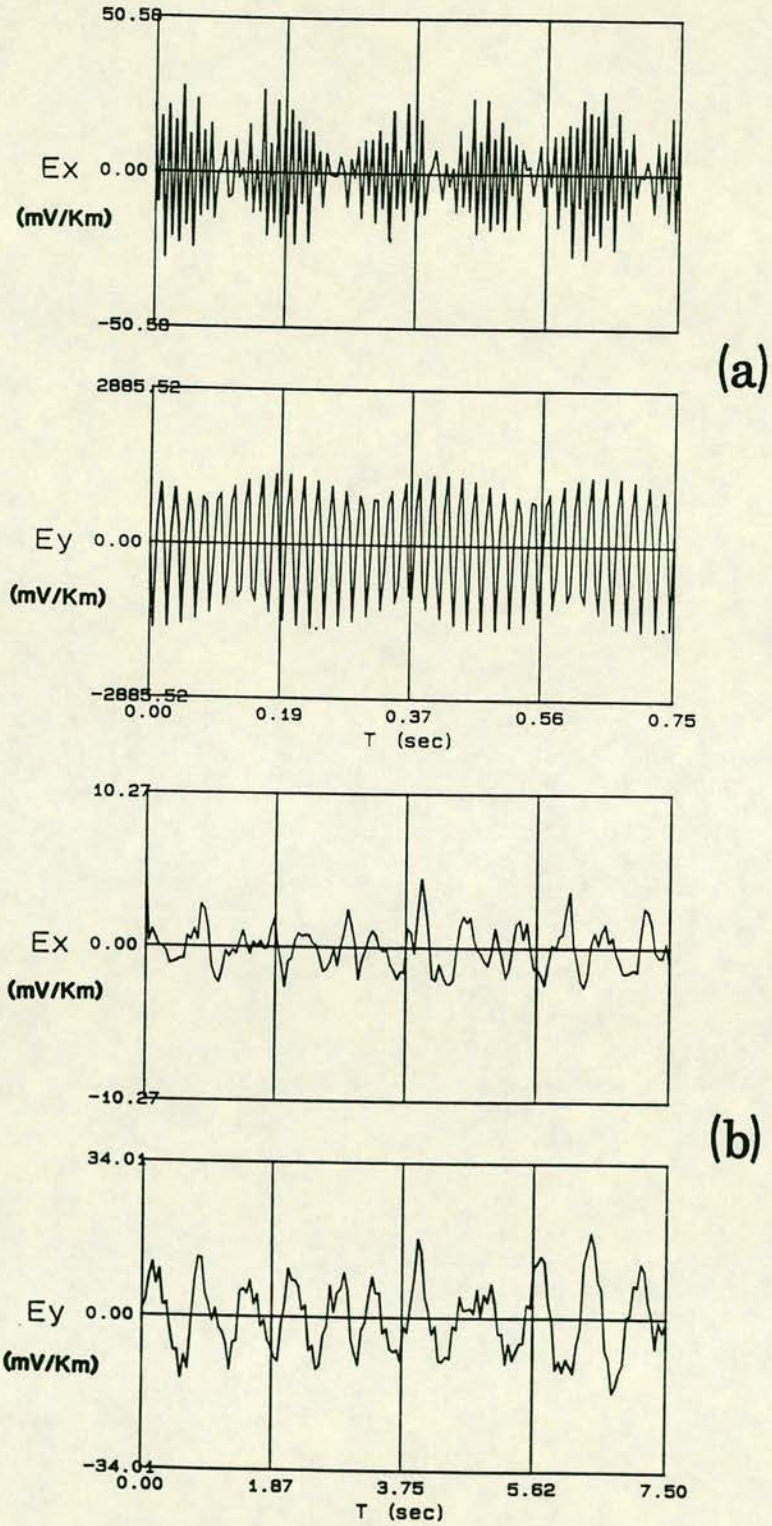


FIGURE 5. Examples of mains noise manifestation near a probable source, for **(a)** decade 1 data, **(b)** decade 2 data. Example site 5.

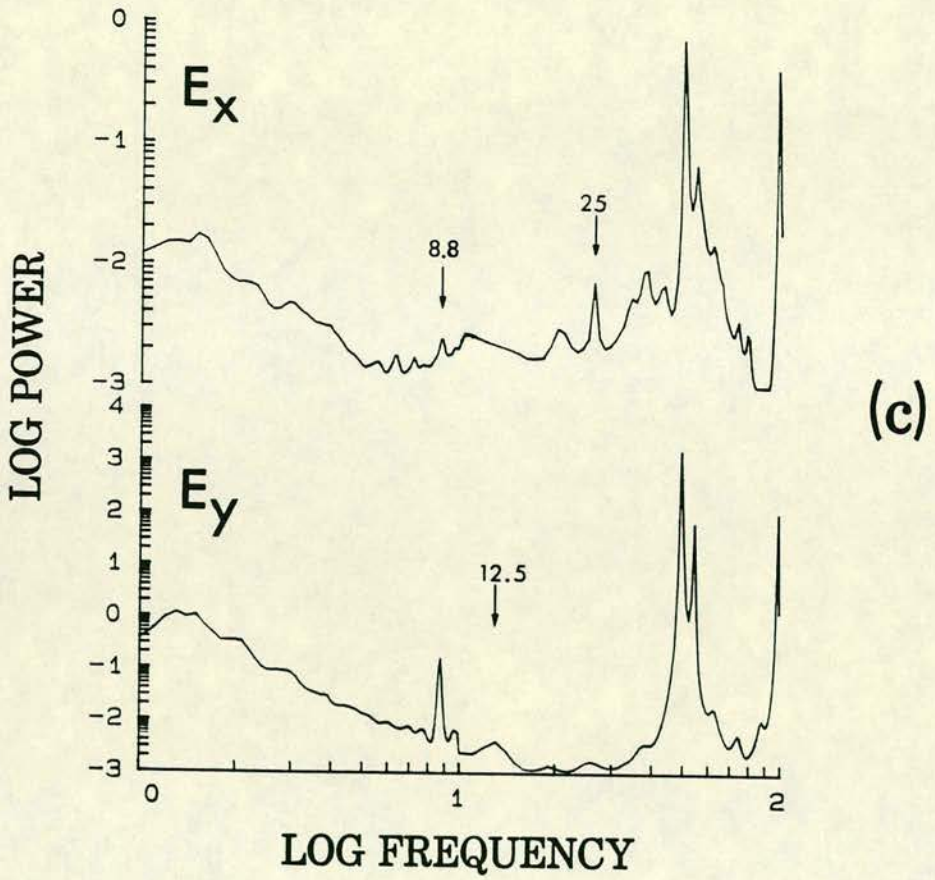


FIGURE 5c. The time averages of the E-field auto-spectra. Decades 1 and 2 of example site 5. Bandwidth extent is 1Hz - 100Hz. Single channel MAXENT spectral analysis of 165 successive data windows for either decade. Data lengths of N=150 and AR operator lengths of M=30 were used throughout.

offers the most probable explanation for 'Feature 1' observed in our earlier presentation of impedance functions.

It may be interesting to note that the results presented above in the form of observed spectra can very easily be simulated with a simple superposition of sinusoids at nominal frequencies of 49Hz and 51Hz in additive noise; thus I have been able to reproduce several of the effects cited above. Having established the possibility of wide band data contamination by narrow band unstable noise processes, let us now examine their 'received' characteristics for site 1. Figure 6 shows the MAXENT spectra for the two telluric (E_x, E_y) and magnetic (H_x, H_y) components for this site. Three decades of spectra are presented, for reasons that will become apparent later. Spectral levels at decade boundaries are necessarily discontinuous. In decade 1, we observe that the most prominent features are the spectral lines at 53Hz and 98Hz, and the considerable reduction in power level at ≈50Hz caused by the instrumental notch. The former lines are again an effect of an unstable power distribution grid, for which the instrumental filter was no match. Apart from these two large scale features, no other directly identifiable noise effect is observed in the power spectrum. The most interesting features can be seen in the spectral structure of the 7Hz-40Hz frequency range. Here we observe a complement of the first 6 Schumann resonance modes (typically 7.8, 14.0, 21.0,27.0,33.5 and 39.5 Hz). They are most clearly seen in the H_x magnetic component. The Q factors associated with all these peaks are of the right order of magnitude (<8); problems however are still abundant, especially in the telluric channels. For the n=1 mode these materialize in the form of an aliased spectral component (indicated with up-arrows), contaminating the 8-10Hz bandwidth. For the higher order modes things are more

complicated. It must be appreciated that at the level of attainable frequency resolution, and, provided that it is sufficiently weak, subharmonic interference could easily be hidden in the resonant spectral peaks. For example, the 25Hz subharmonic would easily be incorporated to the n=4 spectral structure. Likewise, the n=2 (14Hz) resonance, and the 12.5 Hz subharmonic occupy adjacent frequency locations and could readily merge. Thus, the power spectrum appears at first to be consistent. Closer inspection would reveal the existence of small differences in the location of spectral peaks, notably for the n=2 (indicated by broken vertical line) and higher (n=5,6) modes. Also, note that the telluric field spectral slopes are not red, as expected for fields in the Schumann resonances frequency range. Therefore, although individual spectral lines due to noise are not observed, it may be that their influence causes the spectral inconsistency and the ripple in the phase of the response functions for this frequency band.

The lower frequencies in Figure 6, display a stable spectral stucture throughout the whole 0.1Hz - 8Hz frequency range. Here we observe some similarities in the location of some spectral features for the decade 2 and 3 spectra; the numbered arrows indicate peaks that appear to be located in frequency multiples of ten with respect to each other. I cannot account for the existence of these frequency multipliers (or dividers). Spectral peaks that appear consistent over the four data channels, are also indicated with arrows. An appreciable amount of power appears to exist in the 3-7Hz range, particularly in the E_x and H_y channels. Most of the features observed are unaccounted for by theory. I am inclined to label them as man made interference, and I believe that such coherent contaminants may pass the established acceptance tests and produce the kind of distortions we observe.

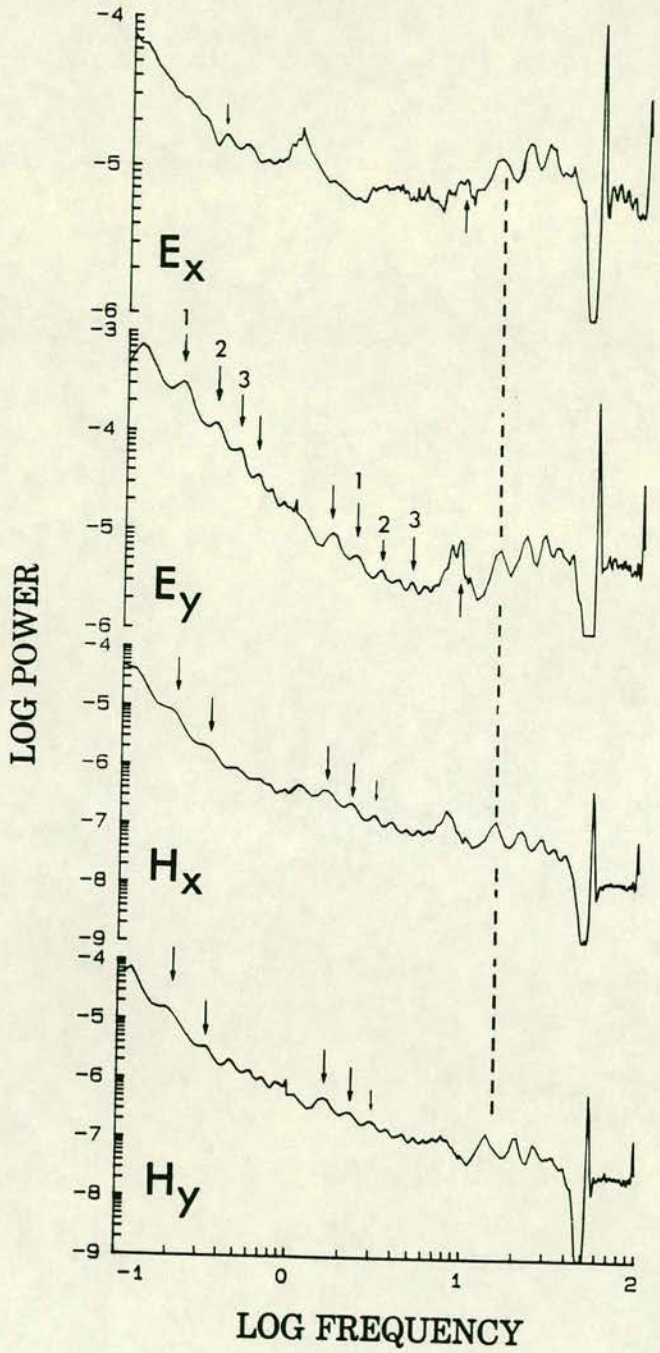


FIGURE 6. Three decades of the time-averaged auto-power spectra of the four horizontal components of the EM field, as recorded at example site 1. Single channel MAXENT spectral analysis of 165 data windows (decades 1 and 2) and 100 data windows (decade 3). Data lengths of N=150 and AR operator lengths of M=30 were used throughout.

VIII. A discussion.

From the results presented above it appears that data adaptive spectral estimation techniques such as MAXENT, when applied to EM data produce transfer functions with properties which are highly localized in frequency. The high resolution affordable by these techniques may be used to improve our insight into some of the processes involving the signal and noise interactions, enabling a better understanding of their behaviour. When the data are contaminated by narrow-band interference, the high resolution afforded by MAXENT appears to enable the extraction of undistorted portions of the bandwidth. Instrumentation cannot always cope with the challenges presented by noise and the requirements of smoothing may integrate very unconforming spectral contributions.

In terms of the contaminated impedance polynomial the main problems encountered were due to structured processes, superimposed on that due to passive natural induction, for which a polynomial representation can always be constructed. The piecewise distortion of the Earth response curve is a result of its multiple convolutions with such noise components. Feature 1 (8-10 Hz) provides an excellent example of such an interaction. It is well known that analog band-pass filters (in this case 8 pole Butterworth), have a maximum delay/phase response. The aliased components, with power at least equivalent to that of the data provide an additional destabilizing agent in the near Nyquist band. The parallel filter rule then demands that the resulting 'noise' wavelet be of maximum phase. The same rule demands that its convolution with the Earth response be at least of mixed phase, albeit with frequency local characteristics. The abrupt departure of the observed phase from its prescribed smooth minimum movement is a clear manifestation of this

effect. A rapid amplitude and phase movement is observed in the data interaction with the mains harmonics and its associated spectral structure. It must be noted here, that a minimum phase wavelet can be extracted by MAXENT, for these components. The parallel filter rule again demands that the poles and zeros of the dominant filter dictate the behaviour of the system. Therefore, upon frequency transformation, the local integration contour is deformed so as to include the poles of the power distribution grid system. The resulting impedance movement may be minimum for the combined system, albeit spurious and unphysical.

Similar arguments can be used to explain noise manifestation in all parts of the spectrum. In such cases, the derivatives and spectral moments of the system, mainly the second derivative $Z''(\omega)$ (rate of change) and second moment (variance) become important as they describe its behaviour. The information conveyed in these functions is valuable when frequency variations are very rapid. One such example is afforded by the results from site 3 (Figure 4a). Let me, however, give another one. In the presentation of sites 1 and 2 no low-frequency decade 3 and 4 response functions have been shown. Decade 4 data were not collected for these sites, and, decade 3 results simply *do not* exist; the noise sources have been formidable for both FFT and MAXENT-based analyses. The MAXENT procedure however has a further advantage in that it can handle short data lengths without significant loss of stability in the spectral estimates (e.g. Tzanis and Beamish 1987b). For the data in question, it turned out that the time duration of the noise sources were much less than that of the recorded data windows. An attractive way to circumnavigate noise problems therefore would appear to be the processing of short time windows. The data were divided so that each N=150 window provided three N=50 reduced length windows. The sample

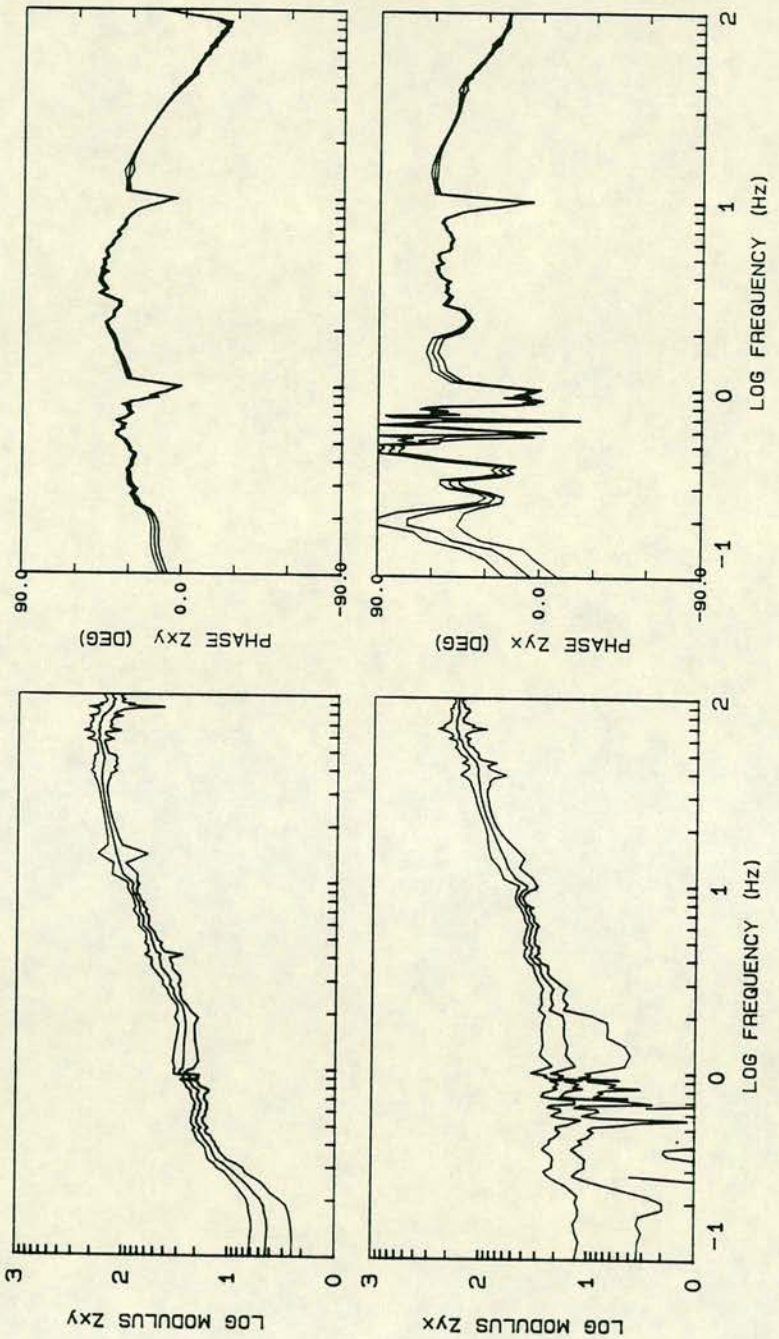


FIGURE 7. The high resolution impedance function for three decades of site 1 data. Decade 1 response is the same as per Figure 2b. Decade 2 and 3 responses are recalculated from reduced length (N=50) data windows, with an AR operator of order M=7.

space, increased by threefold, was then subjected to MAXENT spectral analysis with an AR filter of order 6. The procedure was repeated for decade 2 as well, and the results for site 1 are shown in Figure 7. One can immediately see a significant improvement in the estimation of decade 2, particularly the 1-3Hz band as compared to Figure 2a. The decade 3 response is at last estimated but displays extremely unstable characteristics, with the noise obviously more intense in the Z_{yx} component.

IX. Smoothing a high resolution response function.

The three decade response of Figure 7 provides a good example of an impedance function displaying rapid frequency variation and therefore its alternative description as a high order polynomial can be acknowledged. All the piecewise discontinuous portions imply an unphysical Earth response. They do, however, contain some very important information; they tell us exactly how the contaminated Earth response function behaves locally.

Conventional analysis, by discretizing the observed impedance function discards all such information. As a natural consequence, if smoothing of the response function is attempted it will have to rely on curve fitting of the data with some numerical polynomial interpolation technique, as in Larsen (1975), Hobbs (1982) and others. This however is, at least theoretically, unwarranted, because these methods will fit a polynomial approximation to the data, thus *imposing* on it structure with no a priori information. Parker (1983) objects to the use of conventional polynomial interpolations, arguing that they are

inconsistent with the conditions for the existence of 1D conductivity distributions (i.e. Weidelt's, 1972 inequality constraints). This is quite a strong point, because the interpolating polynomial may not posses the necessary analytical properties, thereby violating causality. This point is also raised by Claerbout (1976, pp. 62), who, however, admits that in many practical situations such approximations can be inconsequential. The only valid interpolations are those that involve the formal expressions for the polynomial Earth response (e.g. Khachay 1978, Parker and Whaler 1981, for the 1D conductivity distribution case). The methods of the above authors are guaranteed to succeed, but require some computational effort, especially in the case of a highly resolved response function. In the case of the impedance matrix polynomial, things are considerably more complicated. I am in agreement with the notion that smoothing procedure should not violate causality. Therefore, I feel that it is important to use the information contained in the observed response, in order to improve on our estimation without imposing some kind of structure. This information can, in general, be extracted from a complete sounding curve and its derivatives.

IX.1 Background to the smoothing procedure.

Consider the function $F(\omega) = u(\omega)\, a^{\omega}$, with the Heaviside step function

$$u(\omega) = \begin{cases} 1, & \omega > 0 \\ 0, & \omega < 0 \end{cases}$$

for real ω. The *observed* Earth response function $Z(\omega)$ can be convolved with $F(\omega)$ to obtain

$$S(\omega) = (2\pi)^{-1} \sum_{n=1}^{\omega_N} Z(\omega + \omega_n)\, F(\omega) \qquad (14)$$

This frequency domain convolution is permissible because $F(\omega)$ is unbounded and continuous over the real ω-axis (i.e. it possesses no singularities), with derivatives and integrals likewise defined, unbounded and continuous. $Z(\omega)$ is analytic over the real ω-axis, and by the Cauchy-Goursat theorem its derivatives are likewise analytic and defined over the real ω-axis. The backward convolution notation in (14) is only taken for convenience. (14) defines a weighted moving summation in the ω-plane; in contrast to conventional averaging techniques, it defines a quasi-continuous operation, adaptive to the constantly varying frequency characteristics of $Z(\omega)$.

Perhaps it is interesting to pause for a moment and consider the integral form

$$S = (2\pi)^{-1}\int_{\omega_0}^{\omega_N} Z(\omega)a^{\omega}\, d\omega \qquad (15)$$

Now, consider that the exponential function a^{ω} is of the standard form

$$dF(\omega)/d\omega = a^{\omega}\ln a, \qquad \int a^{\omega}\, d\omega = a^{\omega}/\ln a.$$

In addition, require that a is such that $a^{\omega} \to 0$ faster, or, at least as fast as $\omega \to \omega_N$. Then integrate (15) by parts to obtain

$$2\pi S = Z(\omega_0)a^{\omega}0/|\ln a| - \int_{\omega_0}^{\omega_N} Z'(\omega)\, a^{\omega}/|\ln a|\, d\omega.$$

It is obvious that the integral part in the above expression can be integrated by parts for a second time to produce

$$2\pi S = Z(\omega_0)\,\frac{a^{\omega_0}}{|\ln a|} + Z'(\omega_0)\,\frac{a^{\omega_0}}{|\ln 2a|} - \int_{\omega_0}^{\omega_N} Z''(\omega)\,\frac{a^{\omega_0}}{|\ln 2a|}\, d\omega.$$

It is apparent that this becomes a recursive operation. The integral remainder can be integrated by parts n times, to obtain, by making use of the property $\ln a \ln b = \ln(a+b)$, the expansion

$$2\pi S = Z(\omega_0)\frac{a^{\omega_0}}{|\ln a|} + \sum_{n-1} Z^{(n-1)}(\omega_0)\frac{a^{\omega_0}}{|\ln(na)|} - \int_{\omega_0}^{\omega_N} Z^{(n)}(\omega)\frac{a^{\omega}}{|\ln(na)|}\, d\omega$$

If we further require that $Z(\omega)$ be of finite order so that its n'th derivative vanishes, the expansion becomes the finite series

$$2\pi S = Z(\omega_0) \frac{a^{\omega_0}}{|\ln a|} + \sum_{n-1} Z^{(n-1)}(\omega_0) \frac{a^{\omega_0}}{|\ln(na)|}$$

Therefore, if there exists an a such that $F(\omega)$ vanishes within the integration path $[\omega_0, \omega_N]$, the summation (14) can be reduced to a series expansion of $Z(\omega)$ in terms of its derivatives, and is finite. The value of S is evaluated on ω_0 , i.e. (15) is simply an integral form of a single step of the complete convolution operation (14), i.e. comprises the output at a single frequency ω_0. Therefore, the value of S, $S(\omega_0)$ can be thought of as the prediction of $Z(\omega_0)$, given a value of a. It follows that although an infinite number of values for a may satisfy the requirement $a^{\omega} \to 0$ in $[\omega_0, \omega_N]$, we are interested in the particular value that mimimizes the error variable

$$X(\omega_0) = Z(\omega_0) - S(\omega_0) \qquad (16)$$

This is a filtering problem, with an easy solution, since there exists only one variable (a) in the 'filter' sequence a^{ω} .

From (16), it is readily seen that

$$X(\omega_0) = Z(\omega_0) - \frac{1}{2\pi} Z(\omega_0) \frac{a^{\omega_0}}{|\ln a|} + \frac{1}{2\pi} Q(\omega_0)$$

where $Q(\omega_0)$ refers to the rest of terms in the expansion. Thus, the 'filtering' operation will actually alter a weighted version of $Z(\omega_0)$, while leaving the rest of the expansion 'unaffected'. One can see this easier, if one considers that the convolution (14) corresponds to a multiplication of the impulse response with a time function of the form

$$f(t) = \left[|\ln a| - (-it) \right] . \left[|\ln a|^2 + t^2 \right]^{-1} \qquad (17)$$

the inverse Fourier transform of $F(\omega)$, for $0 < a < 1$. The factor (-it) is the Fourier *frequency* differentiation operator. The numerator of (17) will damp $Z(\omega)$ by a factor $|\ln a|$ while the (-) sign ensures that the phase remains unchanged during differentiation. Note that the

derivative of a polynomial is a polynomial of a lesser degree, i.e. a *smoother* function. More importantly however, the factor $(t^2 + |\ln a|^2)$ in the denominator of (17) damps the trailing coefficients of the impulse response, thus forcing on it a minimum delay property. Large trailing coefficients are responsible for non-minimum phase variations since their magnitude is related to the effects of maximally dispersed (non-causal) processes. The effect of (17) on the observed noisy Earth response then is a combination of differentiation and weighting that both lead to a smoother function. From the above it can be seen that we are essentially dealing with a time domain process. The crucial weighting function in (17) is left to the time variable t, which, due to the causal properties of the impulse response, will always be relative rather than absolute time.

The frequency domain convolution (14) however, as it stands, is a very ineffective operation. The frequency variable ω may vary within ranges of several decades, and accordingly, a needs to be different for widely separated frequency bands. For example, it must be $a > 1$ if $\omega < 1$, and $a < 1$ if $\omega > 1$. Therefore, (14) cannot be applied easily, and certainly not for values of ω near unity. The problem will be overcome if we take the convolution (14) to be always relative to a starting frequency ω_j, in the same sense that the time variable t is defined to be relative to a zero time t_0, in time domain convolution operations. This can be acheived if we transform (14) in the discrete form

$$S_i = (2\pi)^{-1} \sum_{j=1}^{N} Z_{i-j} F_j \qquad (18)$$

with $F_j = a^j$, $j=1,2,3...$ Then, the condition $a < 1$ ensures that the weight series will decay fairly rapidly, so that only local frequency estimates will be included in the summation. The smoothing operation as applied through (18) becomes universal and can be used to process wide

band as well as narrow band response functions.

The practical implementation of (18) is very simple. A complex version of Robinson's (1967) FOLD subroutine is all that is needed for example. The output will strongly depend on the non-linear properties of the weights a^j, i.e. on the choice of a. If F_j converges too quickly or too slowly, $S(\omega)$ becomes unbalanced tending to vanish in the first case and to overshoot in the second. Thus, the optimum choice of a will lie in a quite narrow range and will be data (noise) dependent; in my experience this is approximately bounded as $0.82 < a < 0.94$. Finally, if an appropriate misfit measure is implemented, the procedure can be automated for the optimum a. For instance, the expression

$$\sigma = \sum_j [|Z_j|^2 - |S_j|^2]$$

was found to work satisfactorily.

IX.2 Applications and examples

I now proceed to apply the technique to noisy field data. Figure 8a shows 3 decades of the *unrotated* off-diagonal tensor elements recorded in a very intense noise environment. Here, the effects of closely spaced noise sources are clearly observed in the extreme variance of the resulting response function, particularly in decades 2 and 3. The net effect can easily be perceived as a random process superimposed on the true impedance *and* the additional systematic perturbations, e.g. Feature 1. The gaps in the near 50Hz components of the sounding curves correspond to frequency estimates that did not pass the acceptance criteria. Observe also that the decade 3 response for this site is more stable than the equivalent for site 1 (Fig. 7). Figure 8b shows the impedance function recovered after only one iteration with a=0.87. Apart from the overall improvement, other detailed effects include the

recovery of the response mismatch near the decade 3 and 2 Nyquists (0.1Hz and 1Hz respectively) and a partial - although not completely satisfactory - reduction of the magnitude of feature 1 in the phase response. It appears that the technique is an effective interpolator. Note however that if long gaps exist within the data, the technique cannot readily cope (e.g. the ca 50Hz range in the Z_{yx} element). Likewise, if large scale inconsistencies exist in the data, the technique should be applied with caution. The effects of large deviations will propagate through the convolution, and may distort neighbouring frequencies. This can be seen, again, in the ca 50Hz range; the convolution can be seen to have affected some of the 35Hz-50Hz estimates. This application has been carried out with Robinson's FOLD routine. In some situations, an explicit coding of the convolution formula (18) may work more effectively.

The following two examples will be for sites 2 and 1. The result for site 2 is presented in Figure 9, after one iteration with a = 0.865. Scrutiny will reveal that the smoothed response is almost identical to the 'correct' asympote of Figure 3b. The results for site 1 are presented in Figure 10a. In this case, the extremely unstable nature of the observed decade 3 function demands a separate initial treatment. This was done with an optimum weight of a = 0.93. Subsequently, the 3 decades were treated as a whole, and the result shown is after 1 iteration, with a = 0.86. In this case as well, the smoothing process is unable to totally remove feature 1, although the reduction of its magnitude is quite dramatic. The 1-3Hz determination of the decade 2 response appears to be in better order. Note however that the reliable recovery of the decade 3 amplitude response is questionable, especially over the 0.1-0.4Hz band. The steep gradients observed in this portion

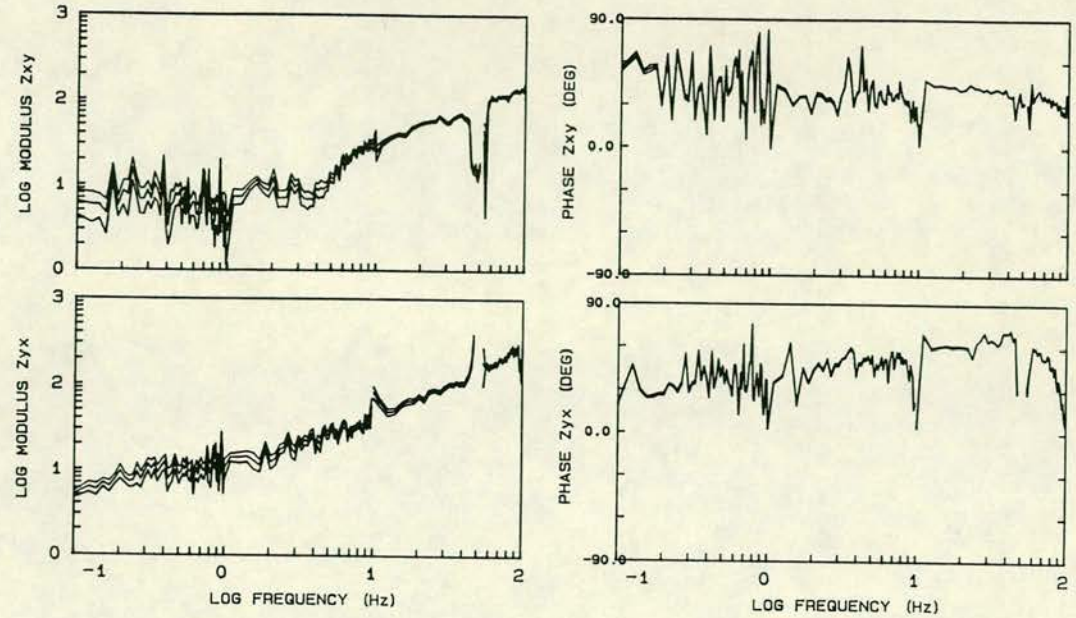
of the bandwidth force the smoothing operator to interpolate and flatten the observation, especially in the Z_{xy} element. With no a priori information about the Earth structure and in the absence of lower frequency impedance determinations, it is impossible to say whether the 'smoothed' function is reliable or not. In view of the noise regime at this site however (e.g. Fig. 6) the noise contamination of this bandwidth appears to be very severe. Moreover, the appearance of local steep gradients in the response function contradicts the intuitively smooth nature of a low degree polynomial response. The same is probably true for the 1-3Hz decade 2 estimates as well. What is important here however, is that the recovered phase response appears to be more reliable for parts of the bandwidth at least. In Figure 10b I present the phase of the impedance 'cleaned' from the residual perturbations and 'conspicuous' bands. If one decides to make use of all available decade 3 information, the remaining phase estimates will provide enough constraints for inversion and modelling.

In view of the above discussion, it worth noting that in cases of severe noise contamination we may be content with the recovery of either a smooth amplitude or phase response function, that the combination of the high resolution estimation and the adaptive smoothing technique facilitates. Not least because either quantity will invert satisfactorily if it is reliable, but also because it may be possible to reconstruct the amplitude of $Z(\omega)$ (or the apparent resistivity) from its phase and vice versa, for they constitute Hilbert transforms of each other. This is a consequence of the causality requirement, mathematically expressed with the Krammers-Kronig dispersion relations (e.g. Weidelt, 1972, Fischer and Schnegg, 1980). Weidelt (1972) gives a method for the recovery of the phase from the resistivity. A very

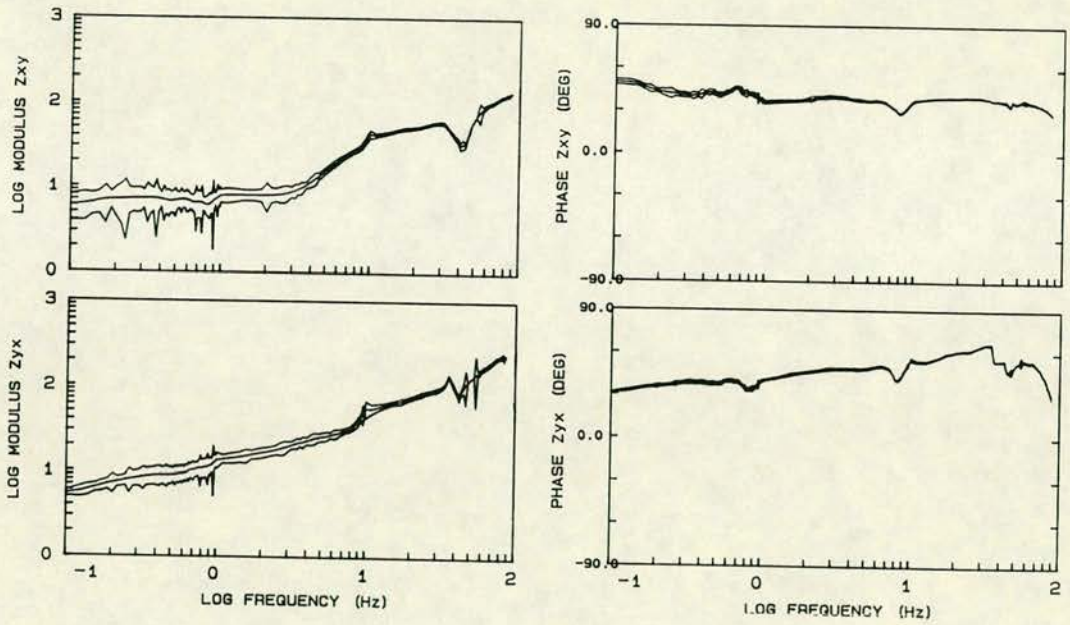
FIGURE 8. **(a)** Three decades of an unstable high resolution impedance function.

(b) The same impedance function after frequency domain smoothing using (18).

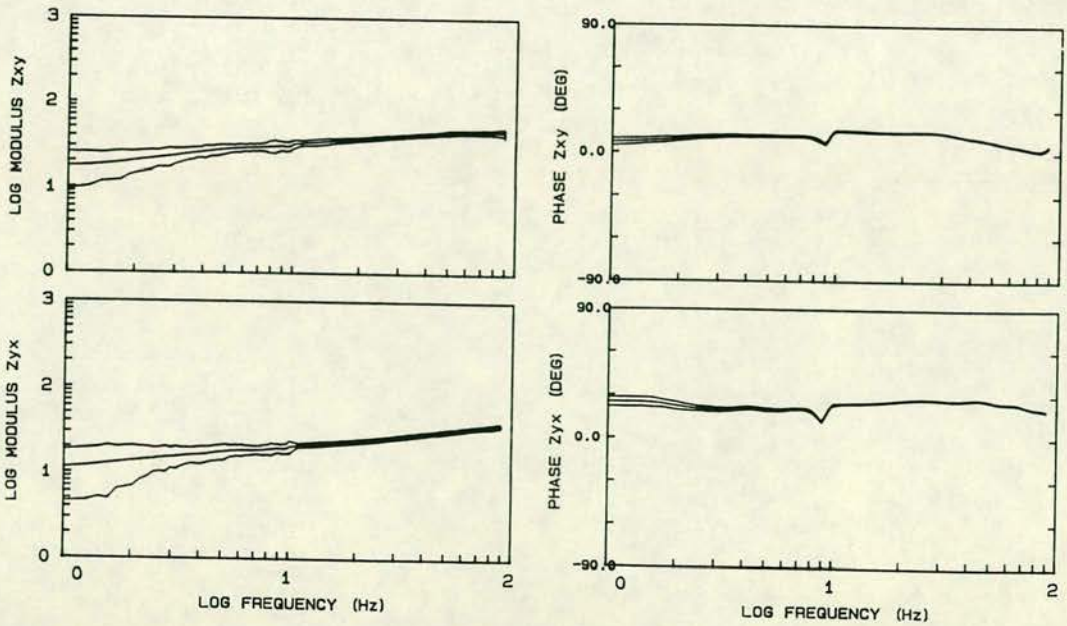
FIGURE 9. The smoothed impedance function of Figure 3a (site 2).

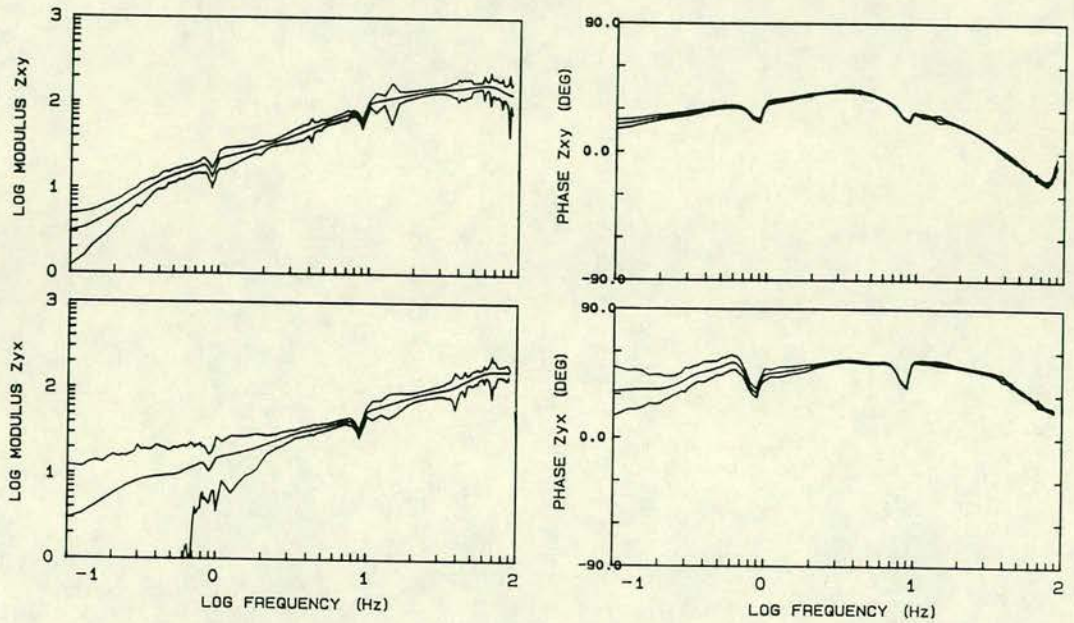

LOG MODULUS Zxy
LOG MODULUS Zyx
PHASE Zxy (DEG)
PHASE Zyx (DEG)
LOG FREQUENCY (Hz)
LOG FREQUENCY (Hz)

(a)


LOG MODULUS Zxy
LOG MODULUS Zyx
PHASE Zxy (DEG)
PHASE Zyx (DEG)
LOG FREQUENCY (Hz)
LOG FREQUENCY (Hz)

(b)


LOG MODULUS Zxy
LOG MODULUS Zyx
PHASE Zxy (DEG)
PHASE Zyx (DEG)
LOG FREQUENCY (Hz)
LOG FREQUENCY (Hz)



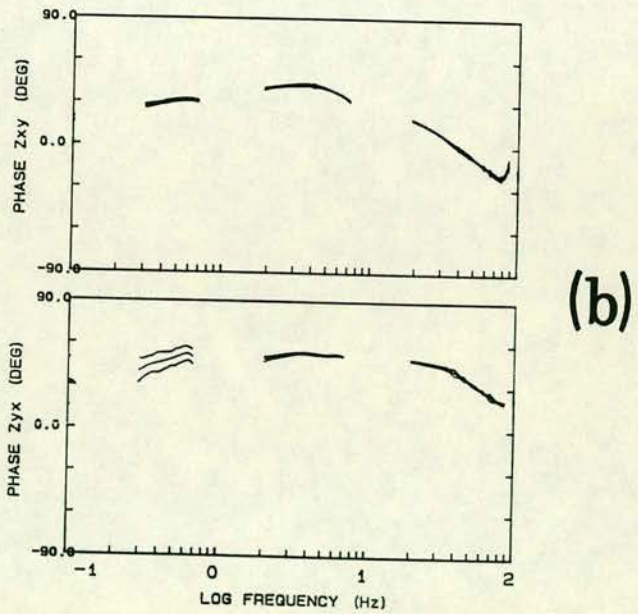


FIGURE 10. **(a)** The smoothed impedance function of Figure 7 (site 1)

(b) The phase of the smoothed impedance function of Figure 10a (site 1), with the remaining unstable and/or conspicuous data segments removed.

interesting algorithm for the calculation of an unbiased *(and also non-statically shifted)* amplitude response from a smooth phase response is given in Clay and Hinich (1981), following the analysis of Boehl et al (1977). In their treatment, resolution plays an important role since the numerical stability of the procedure depends on it.

The interpolating or 'correcting' power of the technique also requires special consideration. The quasi-continuous convolution attempts to to restore a continuous smooth profile along all determinations of $Z(\omega)$. Thus, in Figures (8b,9 and 10a) the smoothed amplitude and/or phase response depart from their estimated position in order to conform to the requirement of continuity. This will be significant if one attempts to invert and interpret the observed and smoothed impedances. Inversion schemes are essentially interpolation processes that minimize some misfit measure between observation and prediction. The effects of the proposed smoothing scheme will therefore tend to constrain the number of ways this interpolation may be realised and thus preempt the resolvable Earth structure; in view of the its properties, it will do so in an optimal way for a given quality of a set of observations.

X. Summary.

In the introduction of the concept of frequency resolution in relation to MT data analysis, I have been primarily concerned with the study and understanding of disturbed AMT observations. Highly resolved impedance functions display their properties in all their glory. Thus, comparisons of spectral features facilitate the correlation of noise effects, not only for single site impedance determinations, but for

regionally collected data as well. These comparisons are at least difficult with the conventional FFT based spectral approach. As it was shown, there may exist cases for which the subtle low intensity noise sources may prove to be more insidious than intense, directly identifiable (and removable in most cases) interference. I believe that such comparisons provide a valuable interpretational aid because, despite being subjective and difficult to quantify, they enable a detailed view of the data quality and what is to be expected of it.

The mathematical properties of a contaminated response function have also been put to work, in order to study a technique that will smooth its highly resolved perturbations. The reasoning I follow stems from the fact that passive EM induction is a continuous process, and arbitrary frequency discretization may sometimes impose conflicting influences on an Earth response function and therefore much of the information it contains may be lost. I have tried to manipulate this information to our advantage with, I hope, some success. Moreover in the example for decade 3 of site 1, the short data handling capacity of MAXENT was implemented, in association with its resolution capacity, in order to extract stable determinations of portions of the impedance function. I believe that such a facility is very important and must be taken into consideration when processing heavily distorted data. However, I must underline the fact that this smoothing method is nothing but an interim solution. It must be appreciated that each time we transform our data, we lose considerable processing capacity, due to the statistical complexity of the frequency domain, and the independence of distinct frequency estimates. The best way to smooth a noisy response function is to *obtain* an optimally smooth response function in the first place. This is difficult, if not impossible, to achieve in the

frequency domain, for data such as those displayed above. An optimally smooth response function can only be estimated in the time domain, through the solution of the filtering problem. There has been considerable progress towards the solution of this problem, and it is hoped that it will be reported elsewhere, in the near future.

Also note, in view of the above discussion, that I *do not* advocate that high resolution MAXENT estimation should become the standard in EM data analysis. The spectral techniques associated with high resolution require considerable computational effort that limits their in-field processing capacity for most existing (high frequency) data aquisition systems. When time is not a critical factor however, and the data appear conspicuous, the implementation of such techniques may assist in the solution of several problems associated with response function estimation.

APPENDIX 2.A Implementation of the multivariate Maximum Entropy Method.

It appears that for the case of spectral estimation from a stationary time series, x(t), there exists a ubiquitous possibility that we can represent it by an autoregressive (AR) process of order M of the form:

$$x(t) = \sum_{m} a(m)x(t-m) + \varepsilon(t)$$

(Ulrych and Bishop, 1975; Jaynes, 1982), where $\varepsilon(t)$ is a white noise error series and a(m) is an absolutely summable filter. Using such a model it is possible to determine the spectrum from the properties of the filter a(m), m=0,..,M that best adapts to the given data set. The power density spectrum will then be given by the expression

$$S(z) = \Delta t.P_m \; / \; (1 - A(z)\; A^{*}(z)),$$

where Δt is the data sampling rate, P_m is the residual error power of the least-squares minimization of $\varepsilon(t)$, and A(z) is the z-transform of the filter. The asterisk denotes complex conjugation. The problem of spectral estimation then reduces to that of determining the optimum filter coefficient vector. The most popular algorithm for estimating the single channel filter coefficients and hence the power spectrum of a stationary time series in the Maximum Entropy sense, is the one proposed by Burg (1968), that eventually became synonymous with the method. In order to apply the spectral technique to impedance estimation however, we require to calculate both auto- and cross-spectral estimates.

Although some techinques do exist for the approximate evaluation of cross-spectra from univariate MAXENT analysis of any two time series (e.g. Ulrych and Jensen, 1974), their application to the calculation of EM response functions will generally return very unstable results. The univariate MAXENT (Burg) AR operator is derived under the requirement

that it is *'maximally non-commital with respect to the unavailable information'* which is the maximum entropy principle of Jaynes (1963, 1968). As 'unavailable information' the principle defines all the information and probability assigments outwith the parameter space under consideration (i.e. the time window under analysis). Theoretically, for a second order stationary process where all the expectation values are time invariant, the AR operator derived from any segment of the time series should be able to process any other future or past portion of the same time series. In practice however, most of the measured data are first order (weakly) stationary, or, only locally stationary. The latter is most certainly true for EM field variations, especially in the Schumann resonances range of frequencies (Beamish and Tzanis, 1986, Tzanis and Beamish, 1987). This implies that there are two things the MAXENT AR operator cannot efficiently do, i.e. (a) predictions (information processing) outwith the time window for which it has been derived, and (b) correlations with other, (even simultaneous) time series, especially when they do *not* belong to the same ensemble, e.g. the telluric and magnetic fields. This is a consequence of the inherent adaptivity of the technique to the data, and the inevitable differences in the information (signal and noise) content of individual data series.

To overcome the above complications, a simultaneous processing of information commonly available in all data channels is needed, i.e. multivariate processing. This can be achieved if we consider a vector time series of the form:

$$\mathbf{x}(t) = [x_1(t)\ x_2(t)\ \ldots\ x_p(t)]^T\ ,\ t=1,\ldots,N,$$

consisting of p simultaneous data channels, i.e. multivariate processing. The MAXENT processor then understands that the data belong to an ensemble with common statistical properties (jointly normal

distribution) and minimises the common noise content in all channels. The equivalent linear AR system will now assume the form:

$$\mathbf{x}(t) = \sum_m \mathbf{a}(m)\, \mathbf{x}(t-m) + \underline{\varepsilon}(t)$$

where $\underline{\varepsilon}(t)$ is a p vector white noise series and $\mathbf{a}(m)$ is a pxp vector absolutely summable filter. The power density spectrum will now be given by

$$\mathbf{S}(z) = \Delta t.[\mathbf{A}(z)^{-1}]^{*T}\, \mathbf{P}_m\, \mathbf{A}(z)^{-1},$$

where $\mathbf{P}_m$ is now the pxp vector residual error power and $\mathbf{A}(z)$ is the pxp vector z-transform of the filter $\mathbf{a}(m)$. The least-squares minimization of $\underline{\varepsilon}(t)$, to provide the optimum unit prediction error filter $\mathbf{a}$, has been considered by a number of authors, (e.g. Strand, 1977; Morf et al., 1978), in a more or less direct generalization of Burg's algorithm. The simultaneous treatment of p data channels provides the opportunity for a direct evaluation of both auto- and cross-spectral components using the pxp operator $\mathbf{a}(m)$.

For EM data consider the vector time series

$$\mathbf{x}(t) = [e_x(t)\ e_y(t)\ h_x(t)\ h_y(t)]^T\ ,\ t=1,\ldots,N,$$

consisting of four data channels. The above spectral procedure can then be used to form the spectral components required for the least-squares solutions. The spectral analysis algorithm implemented here is a modification of the one due to Strand (1977). The modifications did not alter the flow of the algorithm; they were rather aimed at increasing the speed and computational efficiency, especially at the stage of spectral calculations. For the estimation of the random error associated with the resulting impedance tensor elements the analysis according to Pedersen (1982) can be used. In the latter, the statistical significance of the principal component spectral density estimates, is introduced through their associated number of degrees of

freedom; their distribution is assumed to be complex normal for the raw spectral density matrix and complex Wishart for its smoothed equivalent (e.g. Priestley, 1981, pp. 693-701). This is true for spectra calculated with the conventional techniques, however, the multivariate MAXENT spectral estimator lacks an exact statistical description of its properties.

We can circumvent this problem if we consider the statistics of the generalized linear regression system. Its spectral density matrix is shown to comprise a class of consistent, asymptotically unbiased and asymptotically complex normally distributed estimates (e.g. Brillinger, 1981, chapter 8). It is therefore conceivable that the multivariate AR spectral estimator, being a particular case (principal component analysis) of such a system will possess similar statistical properties, although the moments of the distribution are yet to be specified. Such an argument is based on, and enhanced by, the fact that all the entries in the data vector $\mathbf{x}(t)$ are assumed to be second-order stationary time series, jointly normally distributed; this somehow prescribes the result. The number of degrees of freedom associated with the principal components of the spectral density matrix can be taken to be $n=N/M$, as a direct generalization of the result by Kromer (1970), concerning the statistics of the univariate AR spectral estimator. The latter was found to be unbiased, consistent and asymptotically normally distributed. The above arguments provide at the best an approximation, however, they are based on reasonable assessments and are practical with respect to applications in EM field data analysis. A quantitative measure of the goodness of the least-squares solutions for a high resolution (quasi-continuous) impedance function can thus be afforded.

2.A.I On the rationale of MAXENT implementation to EM data - Does it apply?

The question I now address is a philosophical one, however I consider it necessary to bring it up because of the controversial history the method has had with respect to its ability to reproduce the true spectrum of a natural process. The controversy dates from the early days of its introduction as an analytical tool, mainly because of its enthusiastic reception by the scientific community, as a panacea that will deliver spectral analysis from the limitations of the classical techniques. However, problems began to appear which soon led to suspicions of and objections to the method. These can, in general, be classified in the following two categories:

(a) The power spectra returned by the technique may display undesired effects, and

(b) The technique actually fits an AR model to the data, so that it will fail to produce the correct spectrum if the process generating the data is not autoregressive.

The term 'undesired effects' in the MAXENT literature is understood to imply non-physical line splitting and frequency shifting of spectral peaks (e.g. Fougere et al, 1976; Fougere, 1977). Such phenomena have been the subject of rigorous investigations by several authors, as they limit the performance of the method for certain types of data. However, as Jaynes (1982) has shown, such effects are limited to the analysis of 'circular' time series, i.e. ones for which information is repeated at regular time intervals. Thus, the independence of consecutive data segments is reduced by the circularity of the process; this is

simulated as phase jumps that occur at regular time intervals, and gives rise to the appearance of multiple spectral lines close to the nominal frequency. It is therefore anticipated that such problems arise during the analysis of periodic or quasi-periodic phenomena of low order processes, such as for instance are simulated time series of sinusoids in noise with variable initial phases. The fact that such effects occur in the low noise case (Fougere, 1977) and have mainly been observed and investigated with simulated time series, appears to support these ideas. Ironically enough, the adaptivity of MAXENT to a given data set apparently works against it when it is not applied with caution. EM field data, especially in the audio and sub-audio bands, possess none of the above qualities, (e.g. Figure 1), and investigations have shown that such effects do not actually occur.

The second line of objections, concerning the nature of the process generating the data, is to a large extent a theoretical one. It is true that various authors (e.g. Gutowski et al, 1978) have shown that the application of the wrong model to the wrong process will yield an incorrect spectrum. However, most of these investigations are again based on simulated time series generated by pure mathematical processes, Autoregressive, Moving Average (MA), or Autoregressive-Moving Average (ARMA); they study the discrepancies between the resulting spectral estimators with precision, which is neither practical, nor achievable in most real situations. To the best of my knowledge, no straightforward account exists in the literature, of comparisons between processes directly identified from real data. Furthermore, as pointed out by several authors, (e.g. Gersch and Sharpe, 1973; Ulrych and Bishop, 1975, etc), any mathematical ARMA or MA process can be satisfactorilly approximated by a sufficiently long AR sequence. Interestingly enough,

the opposite is not generally true as a MA process is not always invertible. I believe that the best available refutation to such suspicions has been provided by Jaynes (1982), in his treatise of the rationale of Maximum Entropy methods; the author argues for the ubiquitous possibility of AR representation of any physical process. I shall only add that the AR (feedback) model is the only one that can directly determine the unique stable minimum delay polynomial associated with any physical process, and this is exactly what MAXENT does. Beamish and Tzanis (1986) applied MAXENT to the analysis of the time-local behaviour of the Schumann resonances modes. This investigation uses theory for control and shows that MAXENT accurately reproduces the theoretically predicted spectral lines. Armed with this experience I attempted to apply the technique to the estimation of Earth response functions and hope that the reader will find an affirmative answer to the 'Does it apply?' question. I think that, as Jaynes writes, despite the complications, the fact that the AR spectral estimator (or the AR filter operator) is the only one possessing Maximum Entropy properties, is not coincidental.

I would like to close this discussion with a comment concerning the effects of the variation of the order of the AR process (M) fitted to the data. I have consistently used standard lengths of M=10 and M=7 for data lengths of N=150 and N=50 respectively, and one may argue that these may not be the optimum cases for all the data subsets. Although the problem of establishing the optimum linear AR predictor for a given data vector **X** is an important and long standing one in statistical literature, varying M has no significant effects on EM transfer function

estimation because the Earth response is a smooth and continuous function of frequency. I have found that any length M of the order of 10% of the data but no less than 5% will produce comparable results. Empirically a lower limit to M follows from the introductory discussion of the polynomial represenatation of $Z(\omega)$. Thus, it can be said that for the given data bandwidth, the lowest AR order sufficient to resolve $Z(\omega)$ is at least equal to the degree of $Z(\omega)$. Therefore, varying M to very low orders (e.g. less than 4) for short data lengths may be unwarranted. Experience with multivariate AR EM data modelling indicates that AMT data are satisfactorily fitted by processes within the ranges cited above, with a very slow improvement beyond an order of 15% of N. The fitting of higher order AR models to the data may occasionally appear to return better results, but, in my experience it does not improve estimation and the computational expense associated with it is usually not justified. Raising the question of computing efficiency, I have to admit that the use of MAXENT has to be restricted in post-fieldwork processing, unless your field system includes a powerful processor. The temporal computational expense associated with it makes it prohibitive for real-time implementation in small computer systems.

CHAPTER 3

ON THE PROBLEM OF ELIMINATING THE HARMONIC (50Hz) INTERFERENCE FROM THE POWER DISTRIBUTION GRID

In Chapter 2, it became clear that harmonic interference from the power distribution grid can have disastrous effects on the estimation of the Earth response functions, and the considerable improvement observed after its removal. This type of noise problems have been concerning research workers in AMT studies, since the very early days of the method. Some of the known techniques to counter this problem include hardware notch filters, dedicated Search-coil magnetometers with optimum signal to noise (S/N) ratio (e.g. as in Karmann, 1977) and Delay Line Filters (DLF), as in Fischer (1982). Hardware filters are, of course, the most obvious technique, and can be incorporated in the instrumentation. However we have seen that harmonic noise due to an unstable grid can easily evade their action, when some of its harmonics lie outside their operational bandwidth. Self-tuning hardware filters are extremely difficult to create, and have not found any application as yet, at least to my knowledge. The use of dedicated Search-coil magnetometers, with self-tuning capabilities presupposes their availability, and, certain hardware specifications. Truly (in fact most of the times), this not always possible. The software DLF approach appears to be more promising, however it is not problem free. I shall now investigate why.

I. Data cleaning with Delay Line Filtering.

Details and particulars about this method can be found in Fischer (1982). These filters, are in effect differencing operators of the type

$$y(t) = w(t)\,[x(t) - x(t-m)] \qquad (1)$$

where w(t) is a symmetrical window function, x(t) is the recorded (input) contaminated signal, m is the time lag of the delay line, and y(t) is the filtered (output) signal. If the time lag m corresponds to the period of a line harmonic, the effect of (1) is simply to remove the contribution of this harmonic (and the subharmonics that correspond to periods of integer multiples of m, k_m, k=1,2,...) from the output signal y(t). For digitally recorded data, k_m can only correspond to integer multiples of the sampling interval Δt. However, one has to appreciate that the contaminating harmonics may not correspond to any integer multiple of the sampling interval. For example, given a Nyquist frequency F_N = 100Hz, and a sampling rate of Δt = 0.005sec ($1/2F_N$), we have

50Hz ($F_N/2$) is sampled at 0.01 sec

33.3Hz ($F_N/3$) is sampled at 0.015sec

25Hz ($F_N/3$) is sampled at 0.02 sec

20Hz ($F_N/2$) is sampled at 0.025sec

.....

If the noise harmonic lies within one of these sampling intervals, (e.g. 53Hz is sampled at 0.00943sec), the application of (1) cannot remove the whole of its power content, but only part of its sideband structure. For data displaying unstable characteristics such as those at hand, DLF is totally ineffective.

One can improve the performance of the filter by incorporating some kind

of tuning facility in the procedures. 'Tuning' would require the knowledge of the exact frequencies to be removed, and, the establishment of a time delay m suitable for the sampling properties of these frequencies. The first task can be accomplished by a simple frequency transformation (e.g. FFT) and testing for the peak power frequencies, in the vicinity of the appropriate bandwidths; I usually explore the 45Hz-60Hz (band 1), and 90Hz-100Hz (band 2) ranges. It is worthwhile to notice that sometimes, the noise source(s) switches off (or on) during data collection. Thus, we may have a sudden transition from noisy to cleaner data, whereupon the continuation of the filtering process is undesirable. In order to avoid such complications, I usually include a simple detector of excess power, in the aforementioned bands. This tests for the power at peak, detected in, say, band 1 (P_{50}), against the average power detected in the 10Hz-35Hz band, where most of the natural EM field activity (the Schumann resonances) is concentrated (P_S). Alarm is triggered if $P_{50} > nP_S$, where n is a constant that I usually set equal to 1.5; this appears to work quite satisfactorily in most cases. Because of the inherent instability of the noise sources and the data, no prior knowledge of the best values for n is possible. However, a careful post data collection study can always reveal its most useful level. Admittedly, the latter tests can be unnecessary computations, but, in my experience, they have been useful in a number of cases.

Having established the exact peak frequency of the contaminating harmonics, one can proceed to establish the most suitable time delay that can be used to remove it. This task would involve an interpolation procedure, through which the input signal can be resampled at any desired rate. Consider again the 53Hz example, which is sampled at

0.00943sec. An eightfold increase of the sampling rate Δt, for instance, produces a $\Delta t' = 0.000625$ so that $15\Delta t' = 0.009375$ which approximates the sampling properties of the harmonic considerably better than the original interval of $2\Delta t$; the application of (1) with a time delay of $m = 30\Delta t'$, on the resampled input signal, will operate on this harmonic much more efficiently. Alternatively, the 53Hz harmonic can be removed as an odd subharmonic of the $m = 10\Delta t'$ period. Resampling (interpolation) can be performed by various methods; in the examples herein I use a simple linear interpolation between the data points x(t) and x(t+1). The calculation of the desired time delay obeys some very simple rules. Let l be an integer constant such that

$$t = l\ \Delta t'$$

For a sampling rate Δt, the frequency F is sampled at

$$\delta t = F_N\ \Delta t/F.$$

Likewise we have

$$\delta t' = F'_N\ \Delta t/F,$$

where $F'_N = 1/2\Delta t'$. Since we must have $\delta t' = \delta t$, any integer $m' = m/2$ such that $m'\ \Delta t' \simeq \delta t$, will be given by

$$m' = \mathrm{Int}(\ \delta t/\ \Delta t') = \mathrm{Int}(l.F_N/F)$$

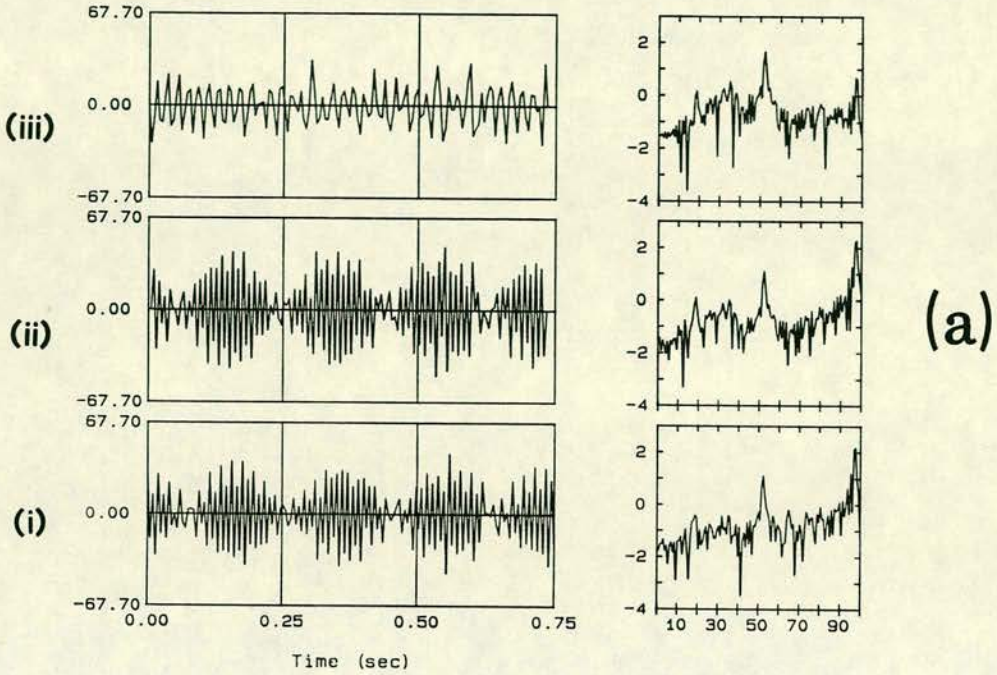
m' corresponds to the best approximation of the half period of a given frequency F, at the sampling rate of $l.\Delta t$.

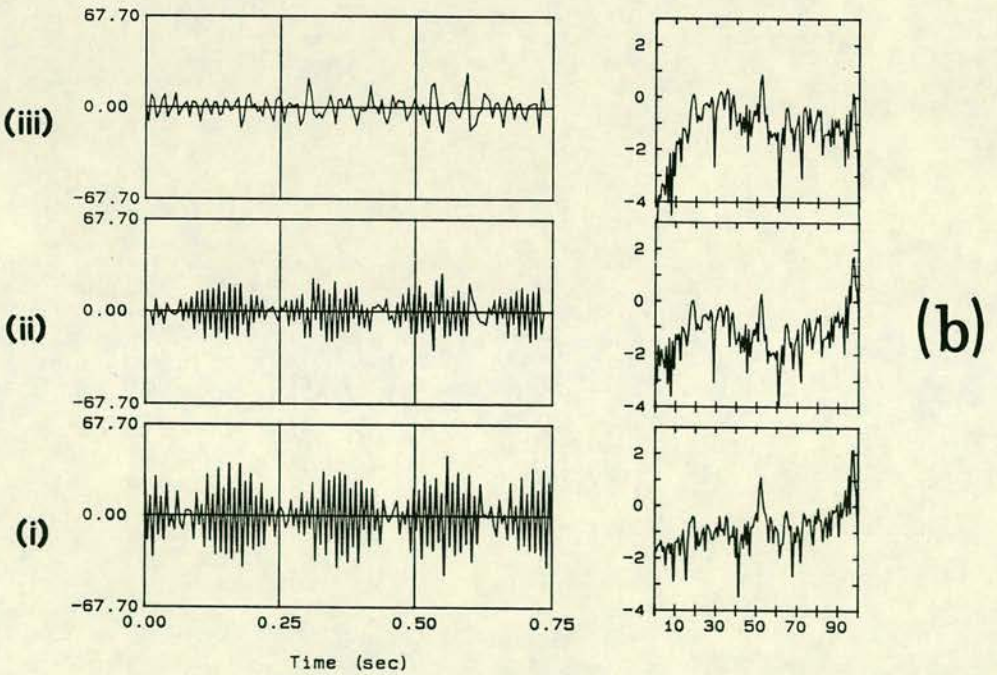
Figure 1 shows the application of this exercise on a severly contaminated data window. The peak frequencies are 52.7Hz, and 97.7Hz. In all figures, (i) depicts the input data series and its associated FFT power spectrum. (ii) shows the result of filtering the 52.7Hz harmonic, and (iii) the final output, after filtering the 97.7Hz harmonic. Figure 1a has been produced by a direct application of (1), without resampling. It is apparent that the 52.7Hz peak remains almost untouched, while

considerable residual noise power exists in the 97.7Hz spectral line as well. Figure 1b is produced after resampling the data at $2\Delta t$. We can see that the 52.7Hz peak is considerably reduced, but the residual noise power still remains at unacceptable levels. The 97.7Hz harmonic is also reduced with respect to Fig. 1a. Finally, Figure 1c is produced after resampling the input data series at $8\Delta t$. Apparently, the 52.7Hz peak is now reduced to acceptable levels, while some difference, albeit not impressive with respect to Fig. 1b, exists in the 97.7Hz peak as well. Resampling at rates over $8\Delta t$ does not appear to improve the performance of the filter, at least for the linear interpolation scheme implemented. I have not experimented with other interpolation techniques. Figures 1d,e provide additional examples from different data sets, after resampling at $8\Delta t$.

Figures 1c,d provide examples of the limitations of DLF when very unstable noise sources are present. Consider that the noise harmonics do not exactly possess line spectra; they are spread spectral peaks. The noise signals experience a great deal of dispersion before they reach our sensors, therefore they should be associated with some sideband structure. Most importantly however, this can be the result of the superposition of multiple sources (e.g. see Chapter 2). DLF can only operate on one line harmonic. If one wishes to wipe out all the noise power, the data must be refiltered, with the danger of inducing gross distortions due to the successive resampling and filtering operations. Another distressing feature of DLF is that resampling at very high rates produces very long data series, and therefore has large storage capacity requirements, while it also becomes a slow process in small computer systems.

FIGURE 1 Examples of the application of the Delay Line Filtering (DLF) method. **(a)** On the recorded (raw) data window; **(b)** after resampling (a) at $2\Delta t$; **(c)** after resampling (a) at $8\Delta t$. **(d)** Is a second example as per Figure 1c. **(e)** Is a third example, as per Figure 1c. In all Figures (i) represents the input (unfiltered) data series, (ii) is the output series filtered for the '50Hz' noise component, and (iii) is the the output series filtered for both the '50Hz' and '100Hz' components.


(iii)
(ii)
(i)
67.70
0.00
-67.70
0.00
0.25
0.50
0.75
Time (sec)
2
0
-2
-4
10 30 50 70 90
(a)


(iii)
(ii)
(i)
67.70
0.00
-67.70
0.00
0.25
0.50
0.75
Time (sec)
2
0
-2
-4
10 30 50 70 90
(b)

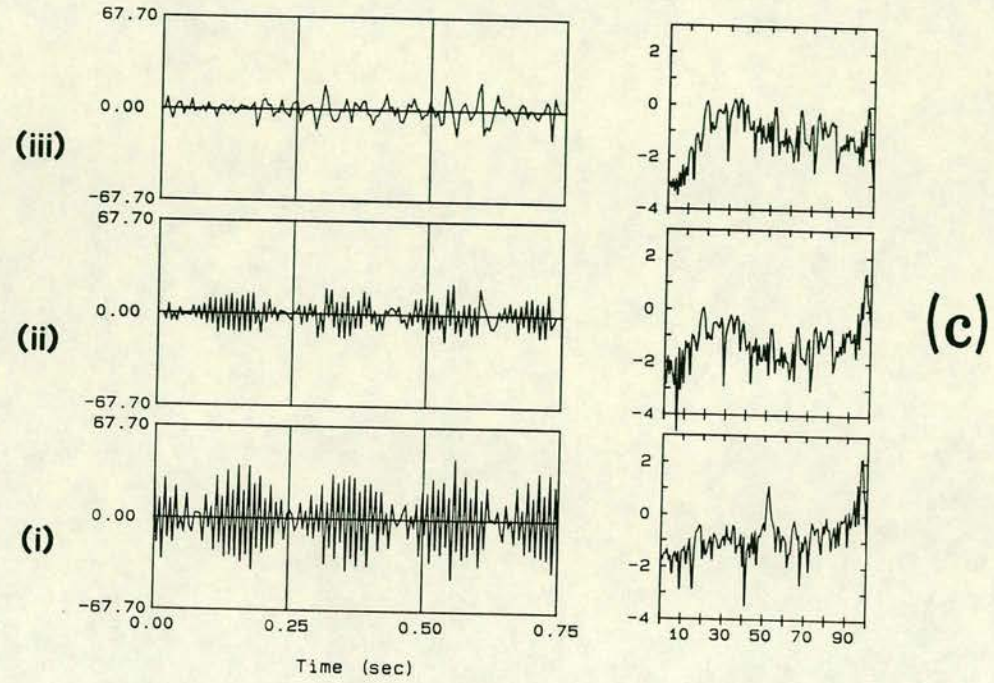
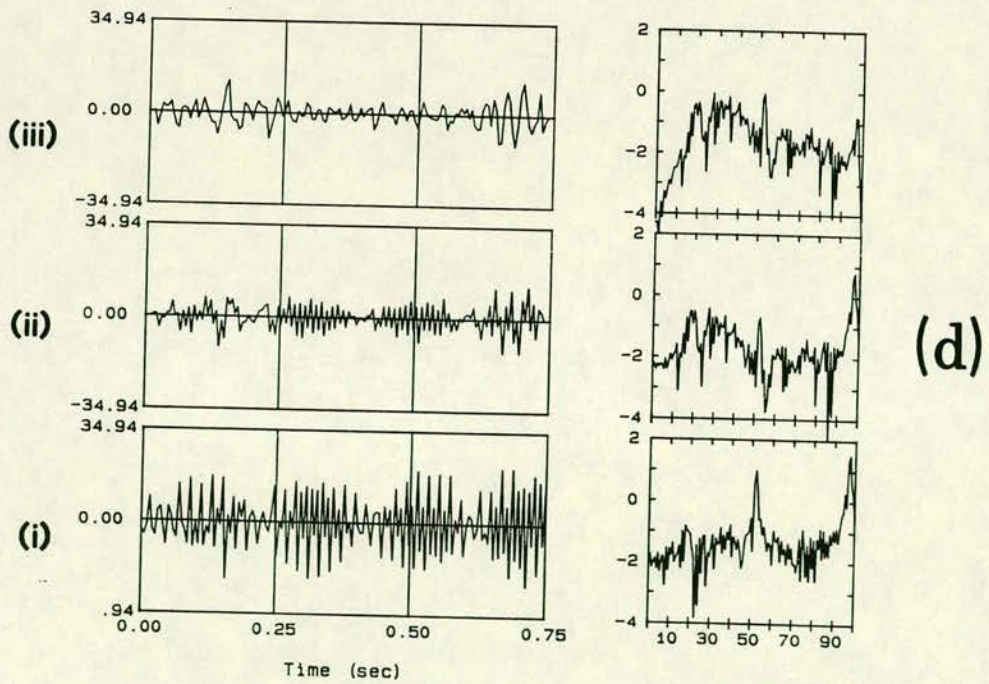


FIGURE 1 ..continued



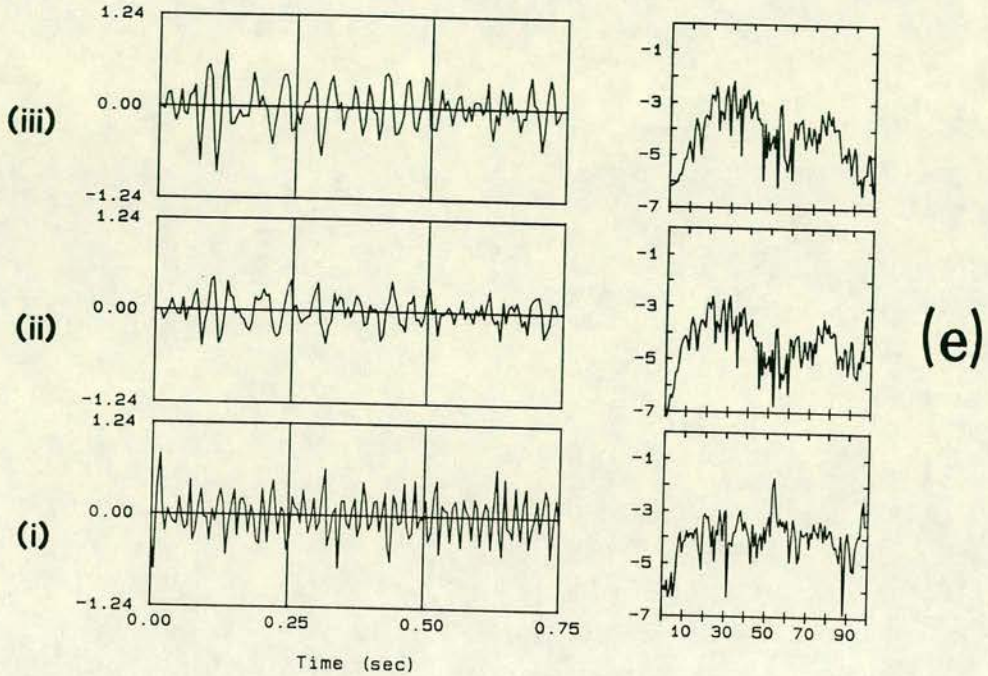


FIGURE 1 ..continued

In order to overcome such problems, for highly contaminated data, I propose an alternative solution to DLF. The philosophy behind this approach stems from the concept of the contaminated impedance polynomial, discussed in Chapter 2. If the contaminating agent is a frequency local polynomial process, then it is possible to model it, and remove the model from the data. Modelling the noise signal amounts to enhancing it, i.e. decreasing the S/N ratio. In this way, one expects that the natural signal contributions will diminish, so that the enhanced noise series will easily be fitted with an Autoregressive (AR), or any other suitable model, that will, subsequently, be *deconvolved* from the data. From the various existing signal enhancement techniques, I have investigated the

II. Data cleaning with Prediction Error Filtering (PEF).

As with DLF, I will not enter into the details and the particulars of PEF, since it is a very well known and publicized method (e.g. Peacock and Treitel 1969, Claerbout 1976, Robinson 1980, and others). In the present, I shall only give an account on the applicability of the technique to our purposes. We can view our contaminated data, as a superposition of two wavetrains, one quasi-harmonic (repetitive) process that represents the noise, and, one uncorrelated random (white noise) series, that represents the Earth's natural EM field. The 'uncorrelated' character of the natural fields must be viewed with respect to the highly correlated repetitive noise waveforms, and their extremely dynamic nature (Beamish and Tzanis 1986, Tzanis and Beamish 1987a). The implementation of a Prediction Error Operator (PEO), aims to separate the highly predictable repetitive time series, from its

uncorrelated and therefore 'unpredictable' part. In general, the PEO with prediction distance α has the form

$$f_{m+\alpha} = 1,0,0,\ldots,0,-f_0,-f_1,-f_2,\ldots,-f_{M-1}$$

where the

$$p_m = f_0, f_1, f_2, \ldots, f_{M-1}$$

part represents a pure prediction operator (PO) of length M. The application of the PO to an input data series x(t), produces an output $\hat{x}(t+\alpha)$ which is an estimate of x(t) at some future time $t+\alpha$. The error series is defined by the difference

$$\varepsilon(t+\alpha) = x(t+\alpha) - \hat{x}(t+\alpha) = x(t+\alpha) - \sum_m p(m)\, x(t-m) \qquad (2)$$

and represents the non-predictable part in x(t). The z-transform of (2) at a time t is (e.g. Peacock and Treitel, 1968)

$$z^{-\alpha} E(z) = z^{-\alpha}X(z) - P(z)X(z) \rightarrow E(z) = X(z)[1 - z^{\alpha}P(z)] = X(z)F(z)$$

The z-transform of the PEO $f_{m+\alpha}$, is simply the difference between the zero-delay unit spike at time t, and the PO P(z), delayed by z^{α} (the prediction distance α). If the prediction distance a corresponds to the period of a highly predictable process, then the PEO will extract from the input X(z) the delayed oscillation $z^{\alpha}P(z)$. Therefore, if one needs to enhance a repetitive process in noise, the problem reduces to that of finding an appropriate length a for its period, and a suitable AR representation for the input (noisy) data series; the latter constitutes the PO p_m.

For the data collection scheme used here, (for more details see Beamish and Tzanis 1986 in Appendix I), one can easily establish the order of a. For a sampling rate of $\Delta t = 0.005$sec, the 50Hz frequency is sampled at $2\Delta t$, and the period of the process is $4\Delta t$. Therefore, at the vicinity of 50Hz, one prediction distance α=4 will adequately describe the oscillatory part of x(t). This would also cover processes in

the vicinity of 100Hz, with approximate period of $2\Delta t$, since these represent the first odd harmonic of 50Hz. Likewise, all the subharmonics with periods equal to integer multiples of $2\Delta t$ are covered by this model. The order M of the PO operator is also of no particular importance; we do not seek to model x(t), but merely to extract a noise model! Therefore, any length that will not overfit the noise process will do. In my experience this is usually of the order of m=12-18. The lower limits are more useful in heavy noise environments, while the upper limits should be used in cases of less intense noise regimes.

Having established the method of modelling noise, we can consider the procedure of removing it from the data. Here, we can run into some problems. The actual prediction error series $\varepsilon(t)$ is not a desirable model for the natural EM field, for three reasons:

(a) The PEO is a phase shifting operator.

(b) The prediction distance $\alpha=4$ is only an approximate, and not an exact model for the noise processes; the difference in equation (2) is subject to the same constraints as the DLF equation (1).

(c) Because the noise power content may vary between data channels. Truly, it is usual that the telluric channels suffer more that the magnetic channels. A consequence of this is that different PO orders are required in order to model the **E** and **H** channels; the resulting differences, however small, produce non-uniform distortion across the data channels, an effect that is probably less desirable than the harmonic noise itself. What we need, therefore, is the 'most objective' description of the noise properties, that does not interfere with the phase, and which can be fearlessly used for all data channels.

Such a facility, can be provided by the unit prediction distance (zero-phase) PEO, that can be extracted from the *modelled* noise series. If we agree that the best noise model can be provided by the most heavily contaminated data channel, so that the uncorrelated part is minimal, then, the best description of its property can be provided by an AR operator that will be extracted from this model. Given that the AR operator is always normalized, i.e. it conveys only spectral shape, and not power information, it can be used to process all four data channels. The cleaned data series will be provided by the prediction error series of this operator. In this way, we can achieve a uniform processor for all our data channels, so that the minimum possible amount of induced distortion due to the filtering procedure is effected on the impedance tensor estimate. The whole procedure amounts to a deconvolution of the noise model from the data. Thus, processing entails

(a) Selection of a data channel x(t) to be used for noise modelling

(b) Derivation of the PO p_m and the noise model $\hat{x}(t)$, such that

$$\hat{x}(t) = x(t) * p_m \tag{3}$$

maximizes the noise content in x(t).

(c) Derivation of the AR operator a_n that best describes the noise model $\hat{x}_a(t)$, so that

$$\hat{x}_a(t) = \hat{x}(t) * a_n \tag{4}$$

(d) Deconvolution of the noise model from the data in the sense

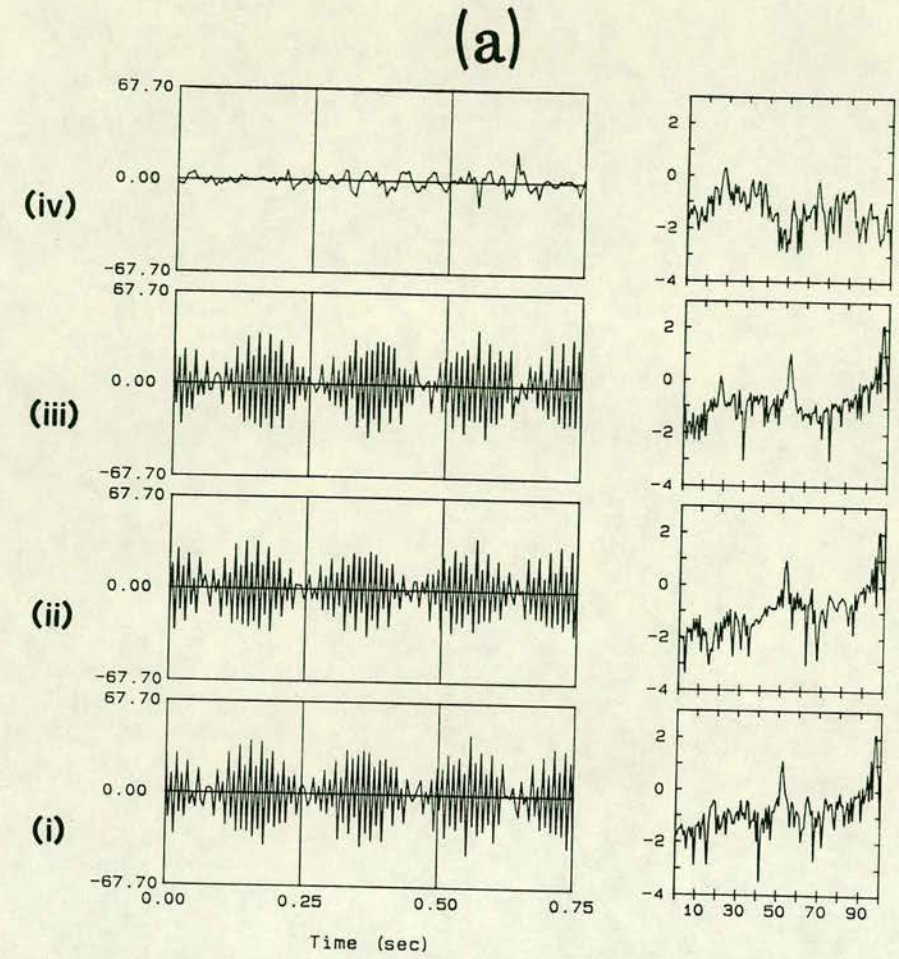
$$d_j(t) = x_j(t) - x_j(t) * a_n \tag{5}$$

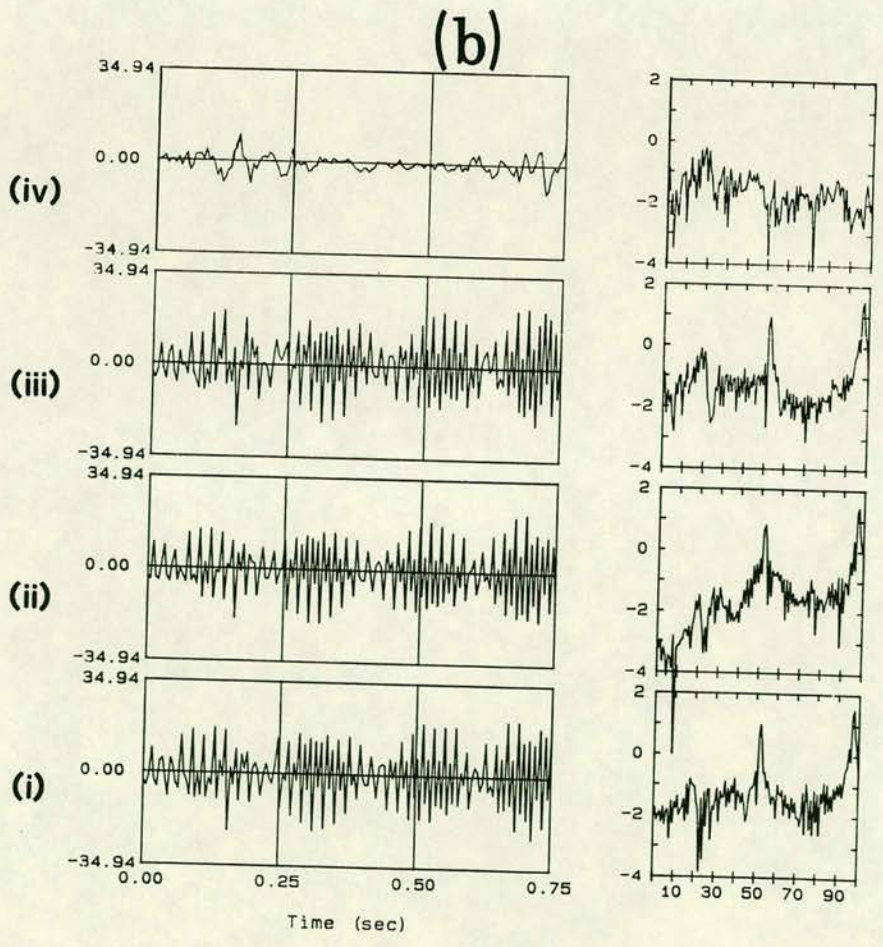
where the index j=1,2,.. runs over the observed data channels. $d_j(t)$ is the cleaned EM data trace; the asterisk denotes convolution. It follows that this procedure can be used in cases of relatively high noise content, so that adequate noise models can be obtained.

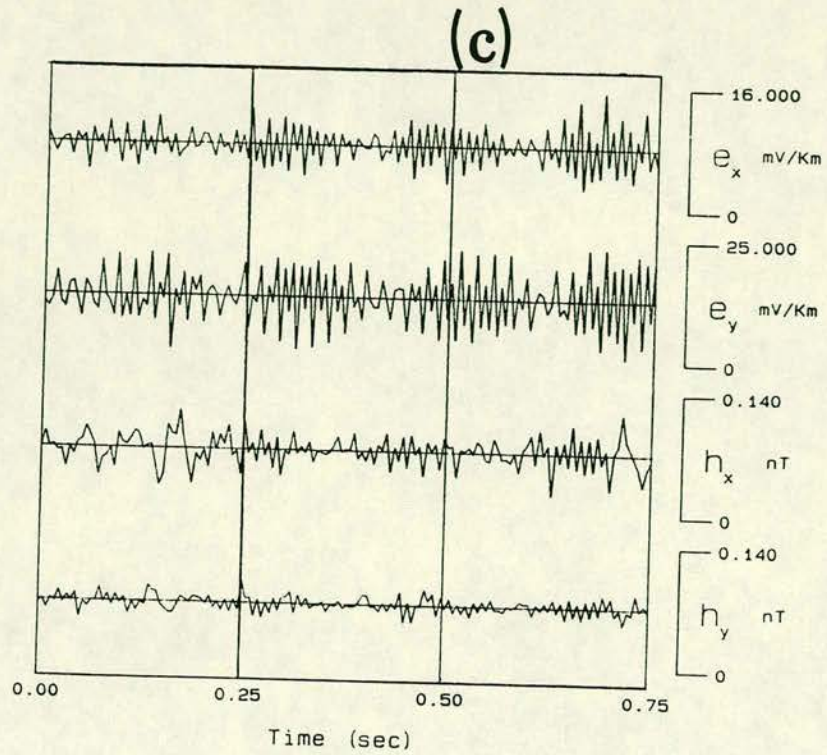
The results of this exercise are shown in Figure 2. In this figure (i) displays the input (unfiltered) data channel, as per Figure 1. (ii) is the noise model, and its FFT power spectrum, as obtained by equation (3). This is used in (4) in order to extract the AR representation a_n of the noise process. (iii) is the result of the convolution of a_n with the input (noisy) data series x(t), i.e. the noise model $x^a(t)$ extracted from the data. Finally, (iv) is the residual (PE) series that represents our clean data trace. Figure 2a is the same data channel, as per Fig. 1a-c. It can be seen immediately, that this procedure radically removes both the 52.7Hz and 97.7Hz harmonics, while it is considerably faster than the DLF resampled at, say $8\Delta t$. Figure 2b is the same data window as per Figure 1d, and provides a further example. Figure 2c,d shows the application of the procedure to a four channel data window. Noise has been modelled on the telluric channel E_y. It is evident, that this procedure (as well as DLF), distort the actual power level of the actual field data (the uncorrelated part of the input series x(t)). Unfortunately, this is an unavoidable byproduct of *any* filtering operation. I will not enter into the details of quantifying the distortion; by applying the same operator to all data channels, one assumes (hopes) that this will be uniformly distributed throughout the data, and therefore minimal on the impedance functions. However, since all kinds of assumptions are optimistic, one should expect to observe some residual distortion in the impedance estimates, induced by the filtering operations.

The PEF method, although effective, is not entirely problem free. Noise modelling relies heavily on the data collection scheme. There may exist some Nyquist sampling rates, that do not facilitate its direct application, if we wish to avoid the resampling exercise. In addition,

FIGURE 2 Three examples of the application of the Prediction Error Filtering (PEF) method. The data series **(a)** and **(b)** correspond to the data series (a) and (d) of Figure 1. In all cases (i) represents the input (unfiltered) data series; (ii) is the enhanced noise *model*, used to obtain AR representation a_n of the noise process. (iii) Is the noise model obtained by convolving a_n with (i); and (iv) represents the residual series of the convolution, which is identified with the cleaned data trace. In all examples of this Figure, the parameters M=12, α=4 have been used for the PE Filter, and n=12 for the AR operator a_n .


(a)
(iv)
(iii)
(ii)
(i)
67.70
0.00
-67.70
0.00
0.25
0.50
0.75
Time (sec)
2
0
-2
-4
10 30 50 70 90


(b)
(iv)
(iii)
(ii)
(i)
34.94
0.00
-34.94
0.00
0.25
0.50
0.75
Time (sec)
2
0
-2
-4
10 30 50 70 90



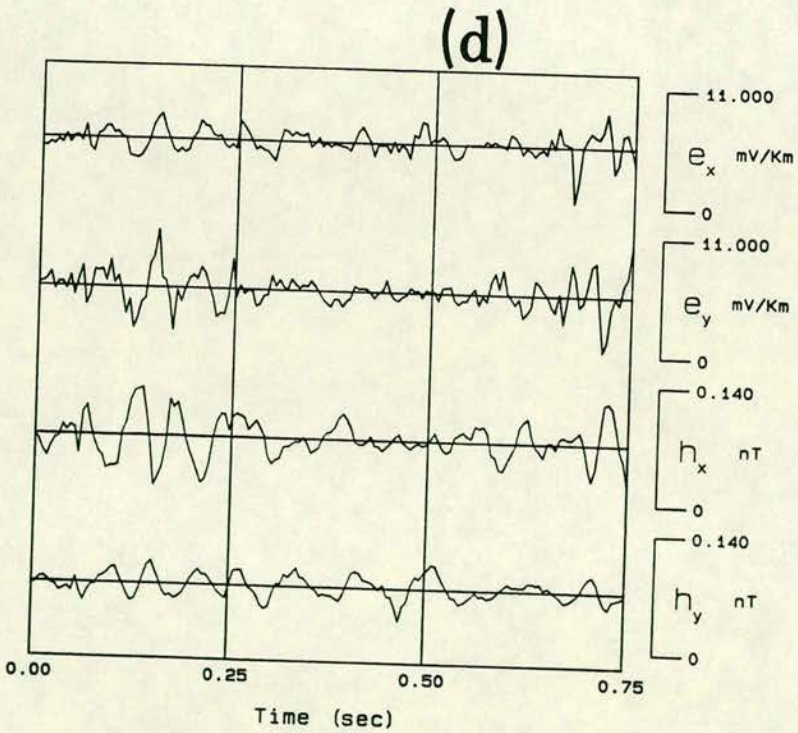


FIGURE 2 **(c)** is a contaminated data window, and **(d)** is the output of the PEF processor; the telluric channel E_y was used for noise mode ling.

one requires an appreciable noise power level in order to obtain a reliable noise model. In cases of low noise content, there exist techniques that can enhance the noise model as an intermediate processing stage (e.g. adaptive filtering, Widrow et al, 1976) However, the addition of other techniques could make the filtering procedure quite complicated and slow. For the above reasons, I seldom make use of the PEF, and rely on a simpler approach that is equally, if not more effective. I now proceed to describe what I call

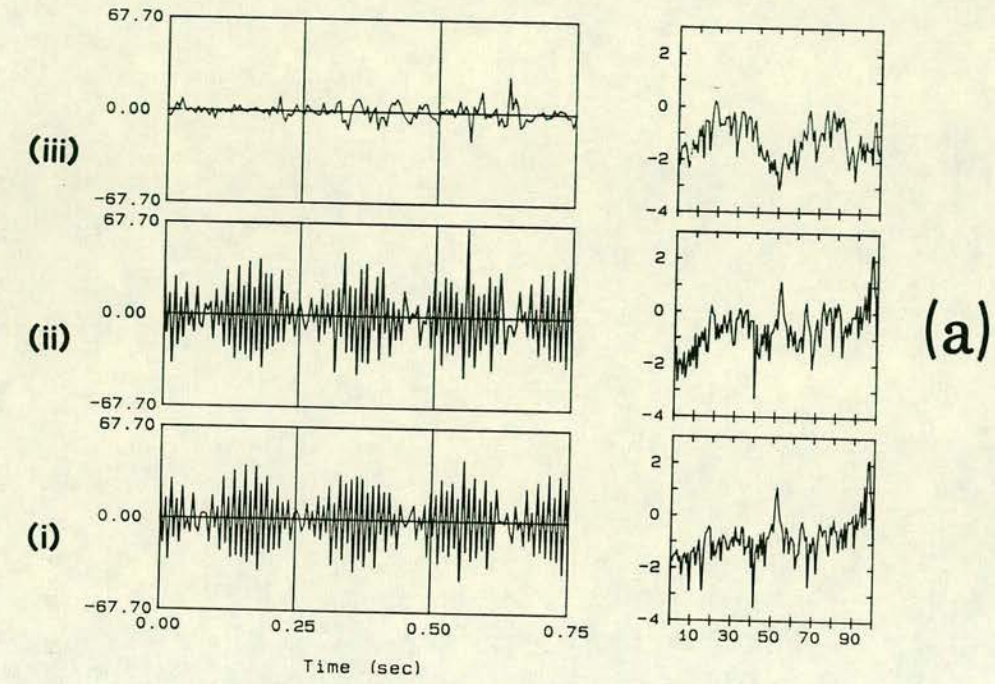
III. Data cleaning with noise simulation and direct deconvolution (SDD)

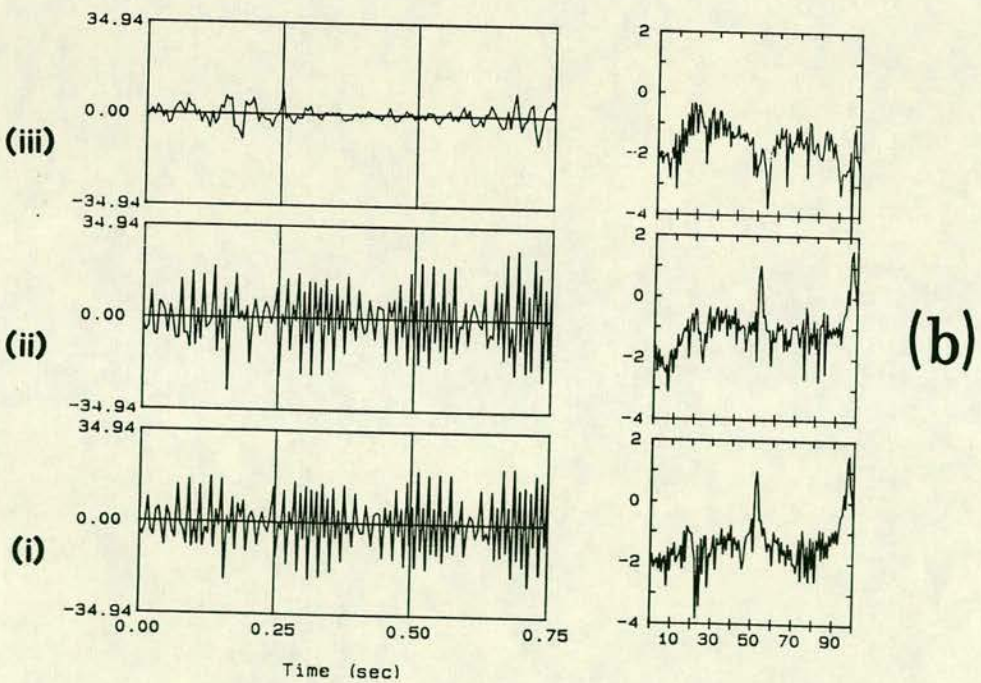
The somewhat impressive title of 'noise simulation', stands for the generation of an artificial time series by superposition of pure sinusoids. As described in I, we are able to detect the peak noise frequencies, *and* their relative power at peak. Therefore, we should be able to utilize this information in order to generate an artificial time series of pure sinusoids, at the detected nominal frequencies, with relative amplitudes (power), proportional to the relative power at peak, of the noise harmonics. A percentage of white noise (up to 20%) can be added, in order to simulate a weakly stochastic process. Next, consider that if no phase information is taken into account, there exists an infinite number of time series realizations that share their autocorrelation function. Furthermore, as the behaviour of the amplitude and phase of one *spectral* component is typically independent of the amplitude and phase of another spectral component within the observable bandwidth, it is possible to ignore phase information and base our filtering on power (autocorrelation) considerations only. The

noise harmonics at peak, and the the zero-phase artificial time series a(t), share the shape of their autocorrelation functions. Our artificial time series becomes, therefore, a noise simulator. One may argue that it is highly improbable (although not impossible) that the peak noise harmonics vary considerably between channels, within the recording time scales, of AMT data windows. Under this assumption, we can use a(t) to compute the *inverse* filter that will deconvolve the noise harmonics from the data. This approach has the basic advantage, that it makes no assumptions about the data properties, and therefore can be very general. Any data channel can be use to detect the noise harmonics at peak. The resulting filter will be a uniform processor for all data channels, thus keeping distortion levels to a minimum. It is worthwhile noting that the length of the inverse filter must resolve, but not overfit, a(t). Therefore, it has to be established with respect to the properties of a(t), (i.e. the number of harmonics to be fitted, and the additive noise content), which can vary, depending on the application. For a limited number of harmonics (up to four) in 20% white noise, I have found that optimum results occur at operator lengths of M=12-20. The additive random noise helps to stabilize the estimation of the inverse filter coefficients. It is recommended that this be carried out with the MAXENT method.

Three examples of the performance of the technique are shown in Figure 3. Here, (i) is the input (unfiltered) data series, (ii) is the noise model (the model that can be matched with the autocorrelation function of a(t)), and (iii) is the residual time series, i.e. the cleaned data. Figure 3a corresponds to the data series of Figures 1a/2a, Figure 3b to the data series of Fig. 1d/2b, and Figure 3c to the series of Fig. 1e. Finally, Figure 3d shows the results of the cleaning

FIGURE 3 Three examples of the application of the SSD method. Data windows **(a)**, **(b)** and **(c)** are the same as per Figure 2 (a) and (b), and as per Figure 1 (a), (d) and (e). In all cases, (i) represents the input (unfiltered) data series, (ii) represents the noise model, and (iii) is the residual series that represents the cleaned data trace. In all applications an AR operator of order L=12 has been used for the representation of the noise simulator a(t). **(d)** displays the output of the SSD processor, for the four channel data window of Figure 2c.


(iii)
(ii)
(i)
67.70
0.00
-67.70
0.00
0.25
0.50
0.75
Time (sec)
2
0
-2
-4
10 30 50 70 90
(a)


(iii)
(ii)
(i)
34.94
0.00
-34.94
0.00
0.25
0.50
0.75
Time (sec)
2
0
-2
-4
10 30 50 70 90
(b)

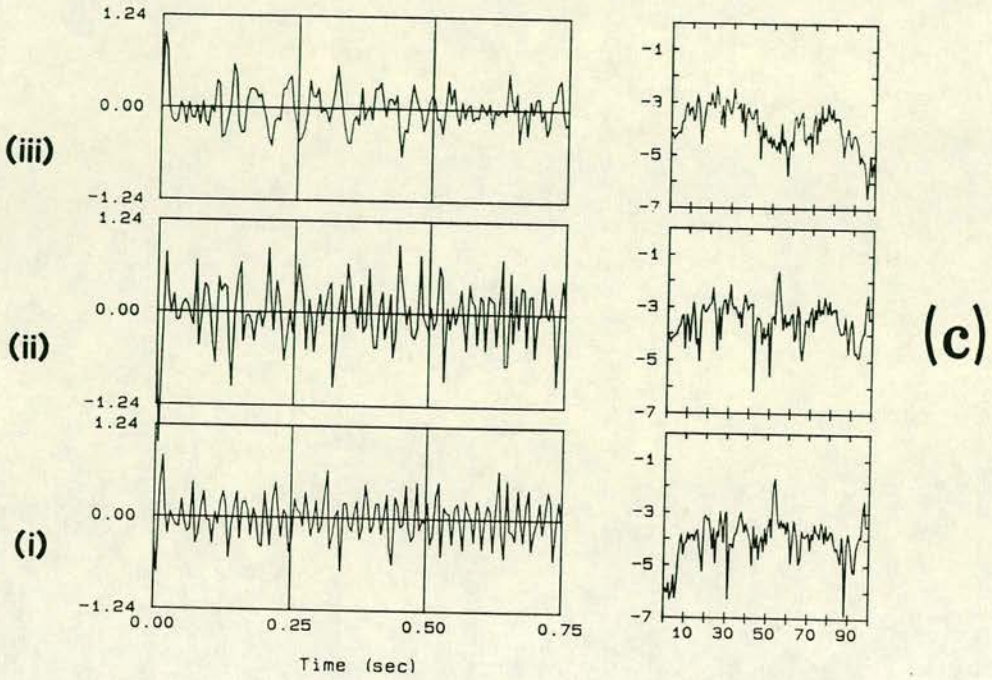
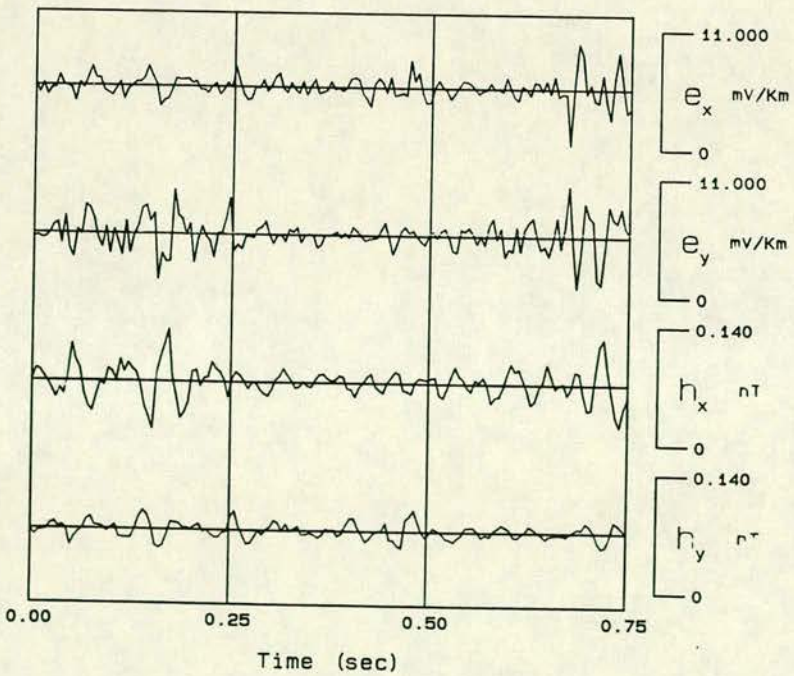


FIGURE 3 ..continued



operation, on the AMT data window of Fig. 2d.

IV. Some general comments.

One reasonable question is that for an assessment of the relative performance of the three techniques described herein. Unfortunately, in my experience, there is no unique answer to this question. Examples of one technique performing better than the other two can be produced, only to be refuted by counterexamples. The problem we face here is that of a combination of dynamic changes (time dependence) in the natural EM field, as well as in the noise sources. The result is a strongly time-local, data (and noise) dependent process. For a stable, well behaved power distribution grid, no problem exists. In the event that any residual noise power is left over by the hardware filters, DLF will remove it satisfactorily. When the grid becomes unstable, the other two techniques appear to give better results, without this being the absolute rule. In very general terms the SSD appears to be the most stable and consistent procedure, while the DLF runs into problems more frequently. The FFT power spectra method displays considerable variance of the raw (unsmoothed) power spectral estimates. This is the main reason why it has been used in the presentation; it can be a very good empirical indicator of the smoothness and stability of a given time series. In the examples of Figures 1-3 it can be seen that, in general, data series cleaned with the PEF and SSD methods display less variance that the corresponding data series cleaned with the DLF operator. This may serve as an indication as to why PEF and SSD may provide more stable filtering procedures in the case of an unstable noise regime. My preference is the SSD, in the event of any indication of instability in

the power distribution grid; data cleaning for the example site 1 of Chapter 2, (as well as Example Site 1, of Chapter 4, has been carried out with this technique. Perhaps it is also worthwhile to mention that the DLF, being a difference operator, amounts to a high pass filter. On the other hand, PO and PEO operators in PEF and SSD act as narrow band, or even notch filters, and leave the low frequency power content unaffected. This will help to explain the differences between the three procedures, observed in the power content of the 0-10Hz band.

What is important however, is that there exists a choice of methods that can be used, one of which may work best for a given data set (and there exist still others to be investigated). Their application and assessment of relative performance for this data set may eventually lead to better estimated Earth response functions.

CHAPTER 4

ON THE PROBLEM OF MT DATA QUALITY EVALUATION, SELECTION AND PROCESSING

I. Introduction

The estimation of EM Earth response functions is a 'statistical' procedure requiring the combination of a number of data subsets or realizations. Given the statistical nature of the estimation procedure we should recognize that the data subsets available may often satisfy the following two conditions:

(a) Some data subsets/realizations may be 'inadequate' with respect to the basic requirements/assumptions of EM induction and/or response function estimation.

(b) Different data subsets/realizations may be mutually inconsistent.

An example of (a) is a data subset containing a contribution (in the magnetic or telluric field) from a current source originating in the Earth. The field relationships cannot then be described by passive EM induction. An example of (b) are data subsets containing different levels of signal to noise (S/N) ratio. The relationships observed between the electric and magnetic fields are then different for the various subsets. In this case the data population statistics (mean, standard deviation) give consistent estimates (expectation values) of the true impedance.

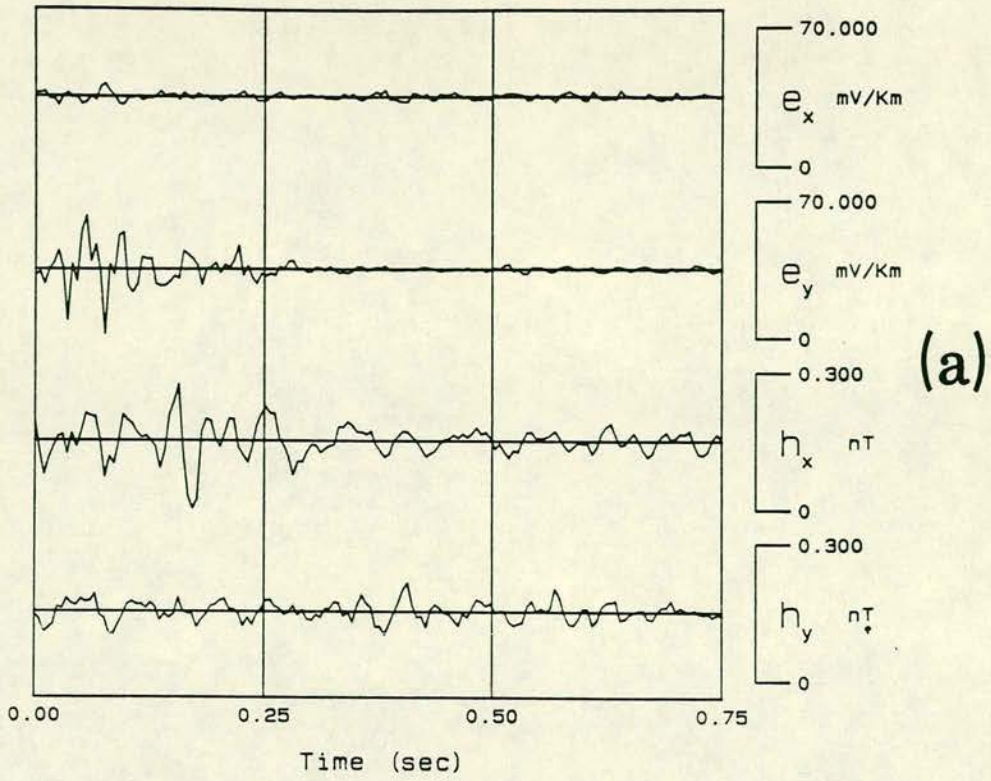
In the ensuing discussion we shall be concerned mainly with case (a) and focus on man-made EM field distortions affecting the audio and sub-audio (Audiomagnetotelluric) frequency ranges. In this frequency range, the

noise sources are due to the power distribution grid and its users. A recent review has been given by Menvielle and Szarka (1986). EM noise from the power distribution grid falls into category (a) above. However, because of its harmonic nature it can be more or less efficiently removed with various methods during in- or post-field processing, e.g. analogue notch filters, delay-line filters (Fischer, 1982), search coil magnetometers e.t.c., and will not concern us here. The noise current waveform pattern however, may be much more complicated than the harmonic structure of the mains emissions. Asynchronous operation of motors can produce additional sidebands and subharmonics of the fundamental mains frequency (e.g. Lokken and Shand, 1964). The switching of current loads can produce both narrow- and broad-band transients and further high frequency harmonics. These transients originate as 'injected' currents and propagate dispersively through the topmost Earth layers, generating secondary (induced) magnetic fields. Their 'received' characteristics are that they arrive at random, and often possess spatially coherent characteristics (stray currents). Such data subsets are inadequate as well as inconsistent, and the usual practices of noise minimization become unwarranted for their treatment.

In Figure 1, four examples of such noise recordings are presented for demonstration. The data used throughout this study span the audio- and sub-audio frequency bands (1Hz-100Hz) and have been described in detail by Beamish and Tzanis (1986, also see Appendix I and Chapter 2). Figure 1a-c displays Decade 1 data (Example Site 1). The natural level of the telluric field activity for this site was of the order of 10-30 mV/Km, and that of the inducing magnetic field of the order of 0.15-0.30nT. We observe short duration (typically less than 0.25sec) current surges associated with an increase in the telluric field amplitudes by any

arbitrary factor ranging from x3 to x12 and, occasionally, as large as x50. These surges are often unidirectional (Figure 1a,b) and are associated with magnetic fields of varying amplitudes, presumably due to the distance from the source, propagation path, and interaction with other sources. Thus, the magnetic field associated with the transient of Figure 1c is of the level of the natural signal, while the equivalent for Figure 1b represents an approximately eightfold increase. Figure 1d is an example of the same effect for decade 2 data, sampled at 20Hz with an effective bandwidth 10Hz-1Hz (Example Site 2). Both sites are located in the vicinity of an industrial area, but quite far from any settlement or apparent industrial activity.

Clearly, such data subsets produce scatter and bias through their coherent contributions towards the statistical estimation of the Earth response functions (the multiple coherent noise of Kröger et al, 1983). In most cases, they invalidate a large number of recorded data windows. It becomes apparent that the basic choice of data length (N) may sometimes play a crucial role in the analysis procedure. Because of the time scales of such noise waveforms, it is worth investigating the possibility of determining the transfer function by analysing very short data lengths, e.g. of the order of N=50. I shall engage in analysis of single events in an attempt to study and understand the time-local and frequency-local behaviour of inconsistent and inadequate data. This will provide the basis for a discussion on possible ways of removing or suppressing the effects of noise, and lead to the description of a simple data selection technique that facilitates the rejection of the heavily contaminated data realisations and their deleterious effects. A comparison of the performance of the most commonly used impedance estimation procedures, under severe noise conditions, is a byproduct of



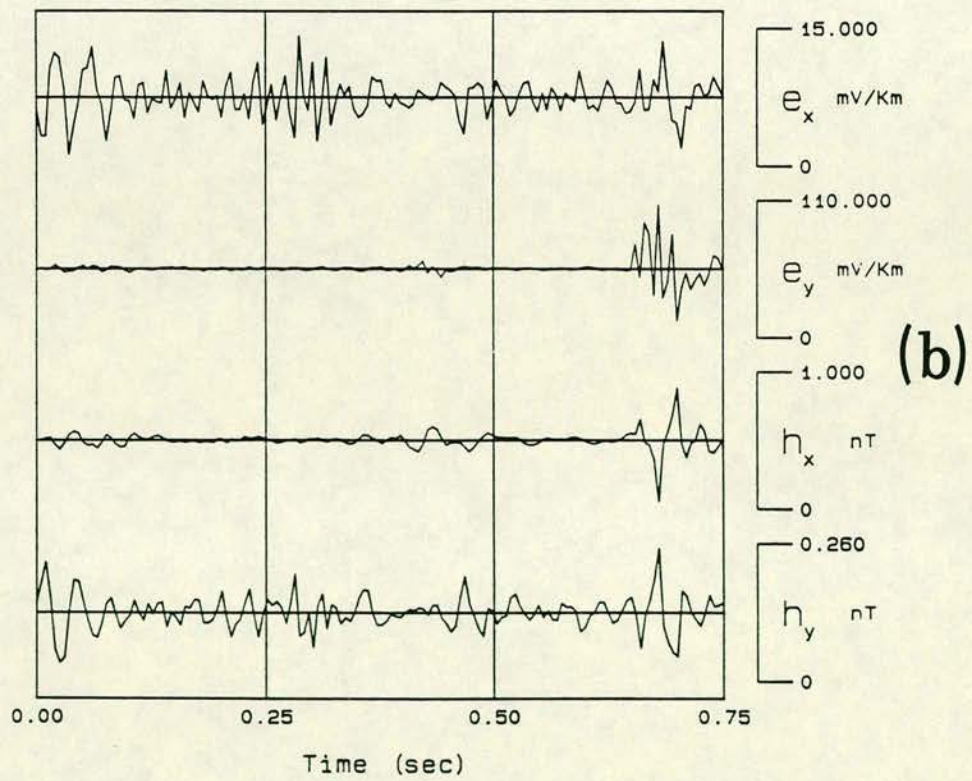
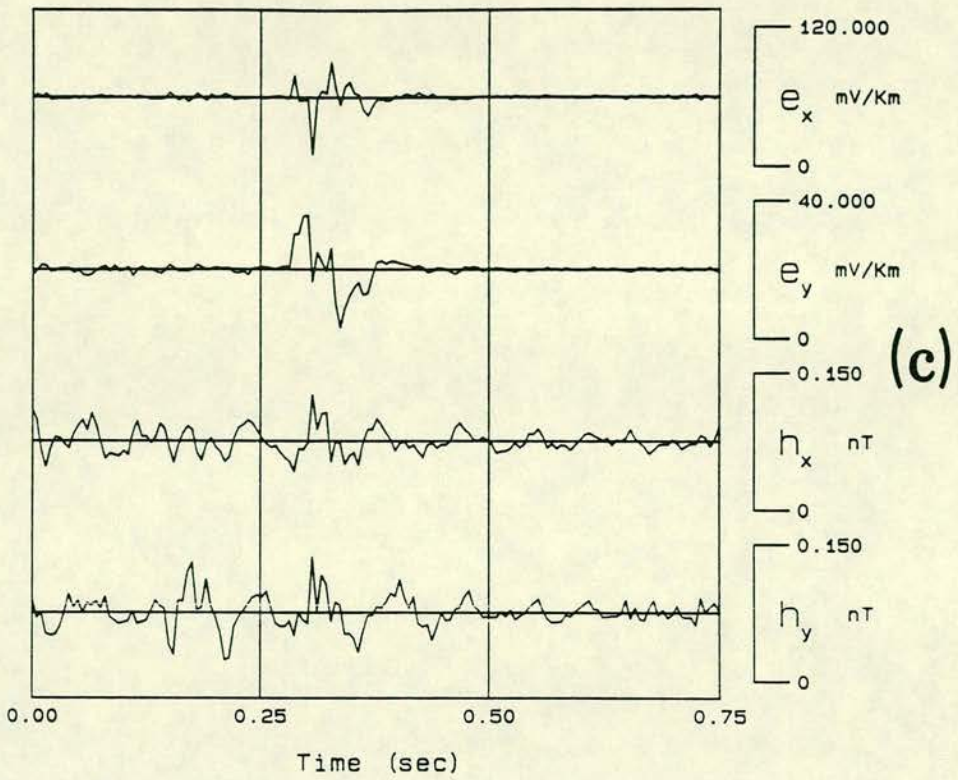
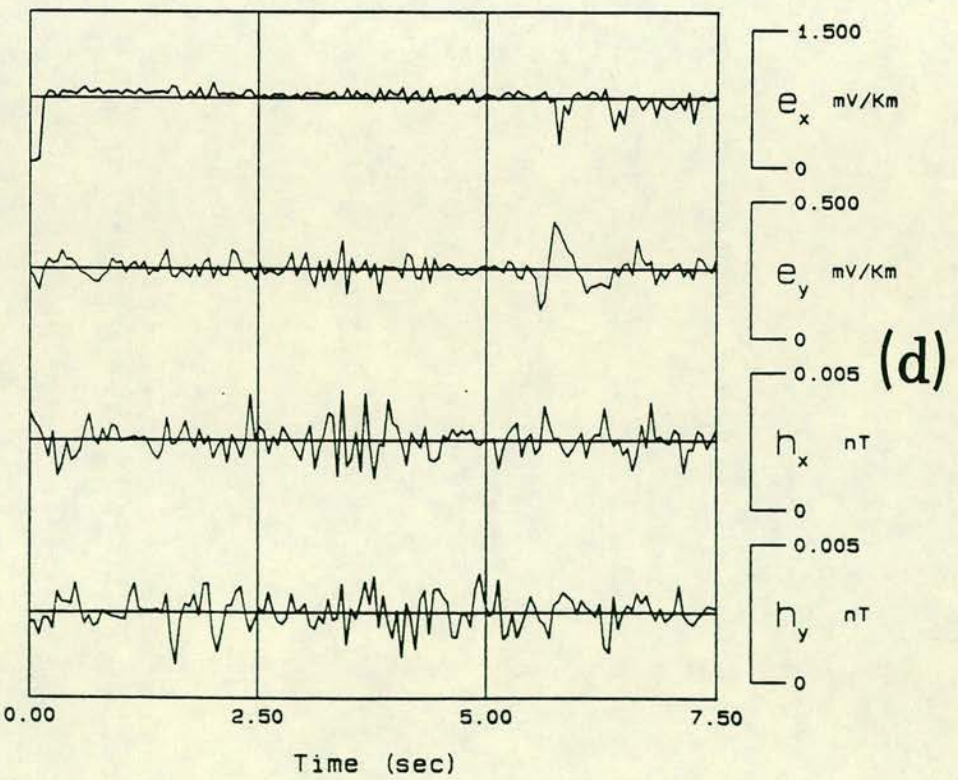


FIGURE 1 4 examples of 'inadequate' data windows, contaminated by powerful industrial transient noise. **(a)–(c)** are decade 1 data from Example site 1; **(d)** displays decade 2 data from example site 2.



(c)



(d)

FIGURE 1 ..continued

these investigations.

II. A spectral study of noise correlation lengths

The conventional FFT spectral analysis method is unsuitable for processing very short data lengths because of the severe reduction in both resolution and statistical significance of the spectral estimator. To overcome this problem, the *multivariate* (multichannel) Maximum Entropy (MAXENT) spectral analysis method has been employed. Implementation, rationale and procedures are described in Chapter 2. In Figure 2a results are produced for the example data window 1a. The amplitude responses of the four downward-biased impedances and their associated predicted coherence functions are shown. Error bounds shown are the 68% confidence limits (Pedersen, 1982). The geoelectric structure in the vicinity of the measurement site appears to be two-dimensional, with the directions of the axes of measurment almost coincident with the principal geoelectric structural directions. In such circumstances, the LS solutions are pathological for the diagonal elements, which become very unstable. Consequently I shall consider only the off-diagonal elements. The diagonals are included in the presentation as well, because of the striking differences in the quality of estimation that can be observed. In Figure 2a the impedances have been computed by analysing the whole (N=150, M=10) data window. Figure 2b shows the equivalent result using the second half of the same data window ($0.38 < t < 0.75$sec, N=75, M=7), i.e. omitting the transient. The difference is striking. Although I shall not attempt to describe the complex pattern of data and noise coupling that emerges for this data window, the bias coupling between Z_{yx} and Z_{yy} (Figure 2a)

caused by the 'injected' industrial transient noise is notable.

In Figure 3, the equivalent results for the example window 1b are produced. Figure 3a is the response computed from the whole (N=150, M=10) window; Figures 3b and c are the equivalent for the 0-0.25sec and 0.25-0.50sec data portions respectively, with N=50 and M=6 in both cases. By comparing the results obtained over the three time intervals, we observe the effect that the noise transient has on the Z_{yx} and Z_{yy} impedance elements, and especially the complex signal and noise interaction that produce downward bias and scatter for the low frequencies and upward bias and bias coupling for the high. The high predicted coherence associated with some of these biased frequency components could ultimately cause them to be included in the population from which the impedance tensor will be estimated.

By studying the results in Figures 3b,c yet another interesting pattern emerges. If we first concentrate on the Z_{yx}, Z_{yy} tensor elements, observe the profoundly different levels of the S/N ratio encountered within 0.50 seconds, i.e. the very high coherences associated with the 28-70Hz band for the first 0.25sec of the window (Figure 3b), and the exact opposite for the following 0.25sec (Figure 3c). Note also the frequency-local, low coherence contributions at 20Hz for the 0-0.25sec interval and the relative improvement for the 0.25-0.50 interval. Equally interesting is the study of the complex destructive interference pattern that produces the response obtained from the whole data window (Figure 3a).

The study of the the Z_{xx} and Z_{xy} tensor elements in Figure 3, provide provide equivalent insights. The analysis of the whole data window here produces results unaffected by the unidirectional

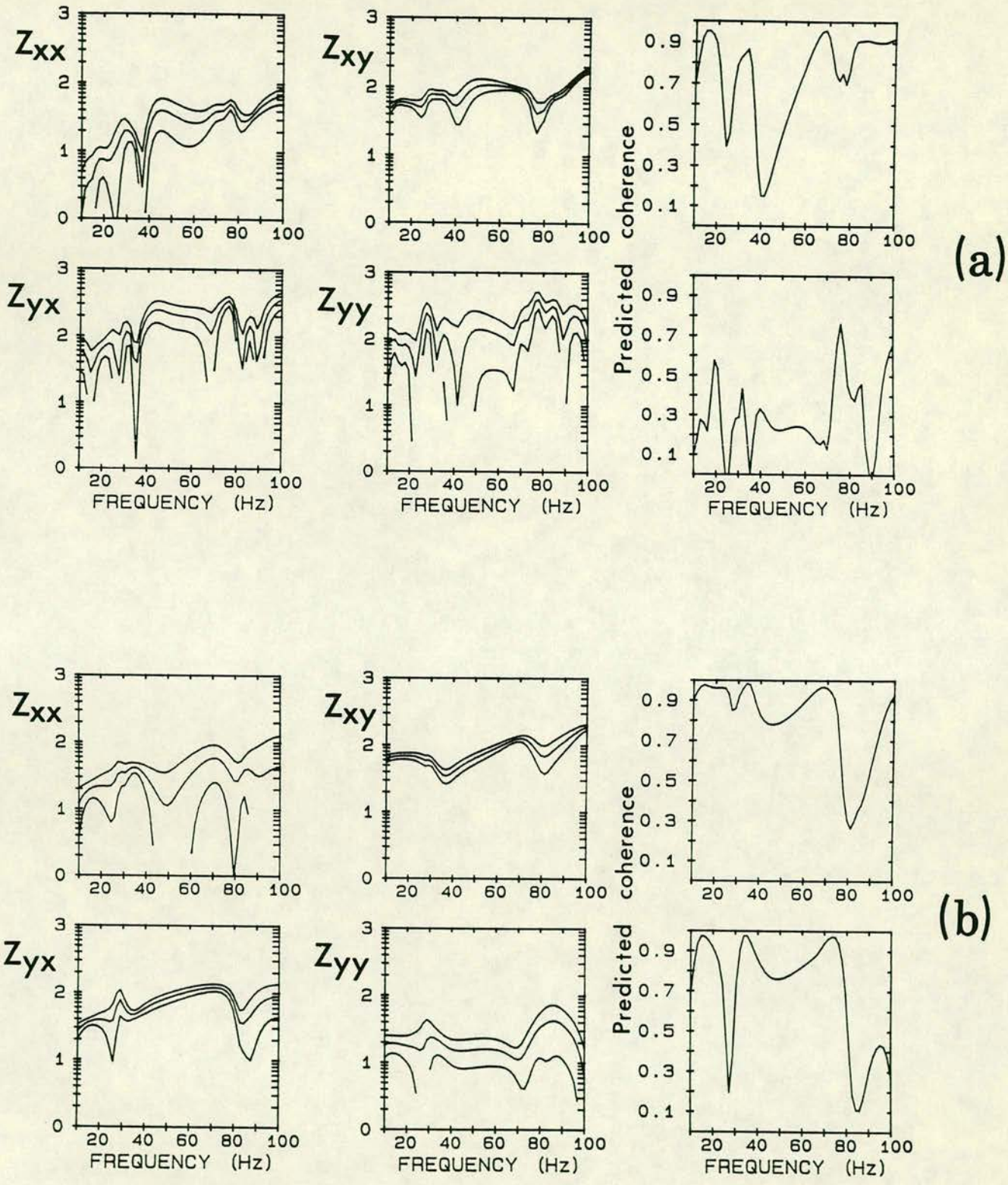
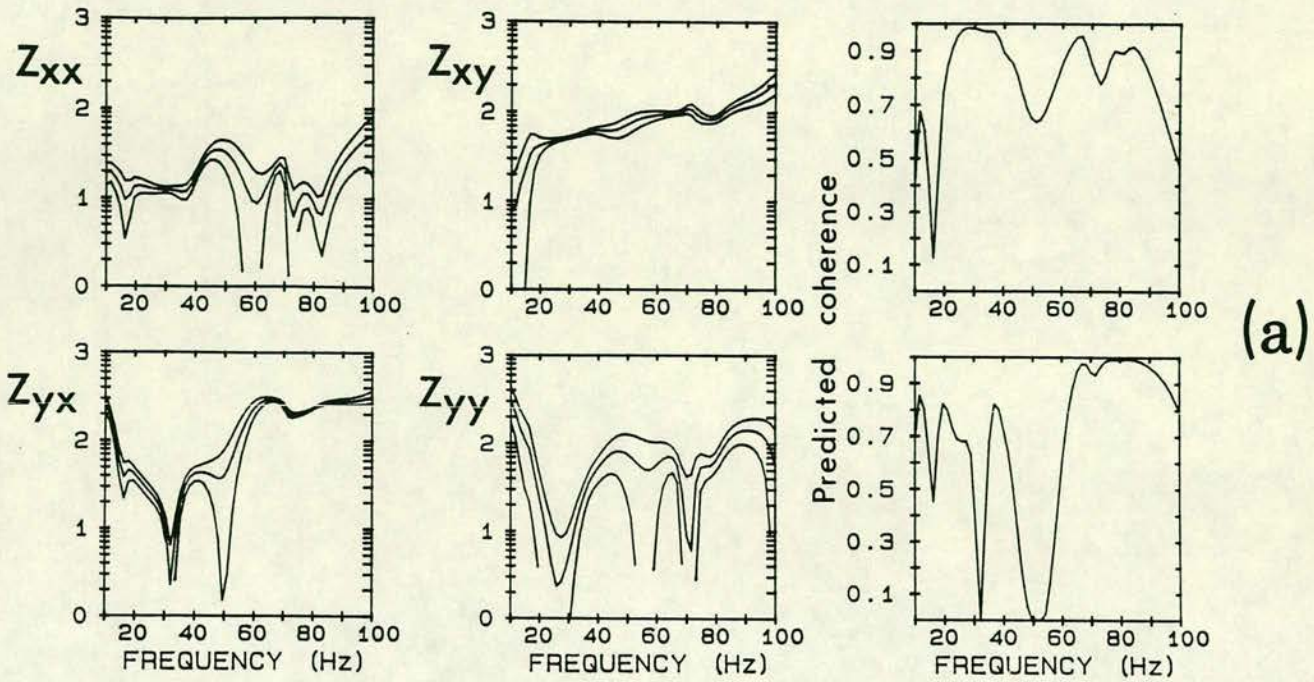
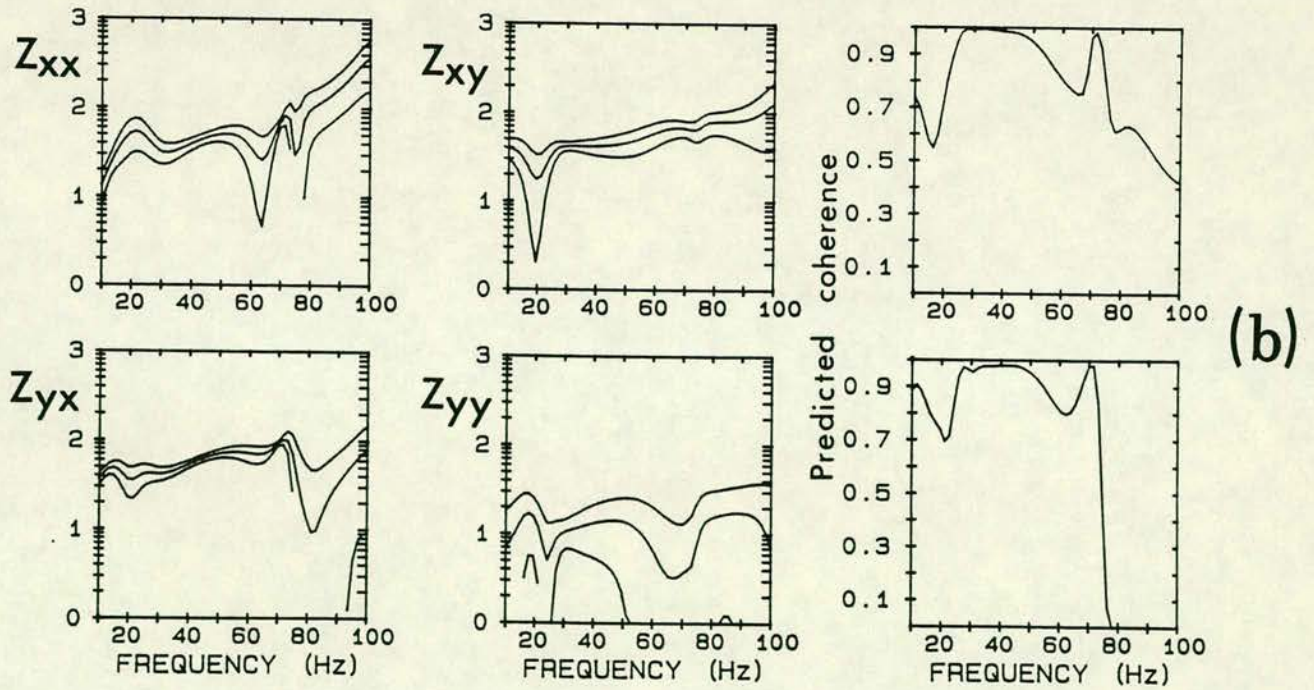


FIGURE 2 Analysis of the Figure 1a data. **(a)** The amplitude response computed from the whole data window (N=150, M=10). **(b)** As per (a) for the second half of the window (0.38<t<0.75sec, N=75, M=7). The parameter M refers to the order of the MAXENT AutoRegressive operator extracted from the data.


Zxx
Zxy
coherence
Zyx
Zyy
Predicted
FREQUENCY (Hz)
(a)


Zxx
Zxy
coherence
Zyx
Zyy
Predicted
FREQUENCY (Hz)
(b)

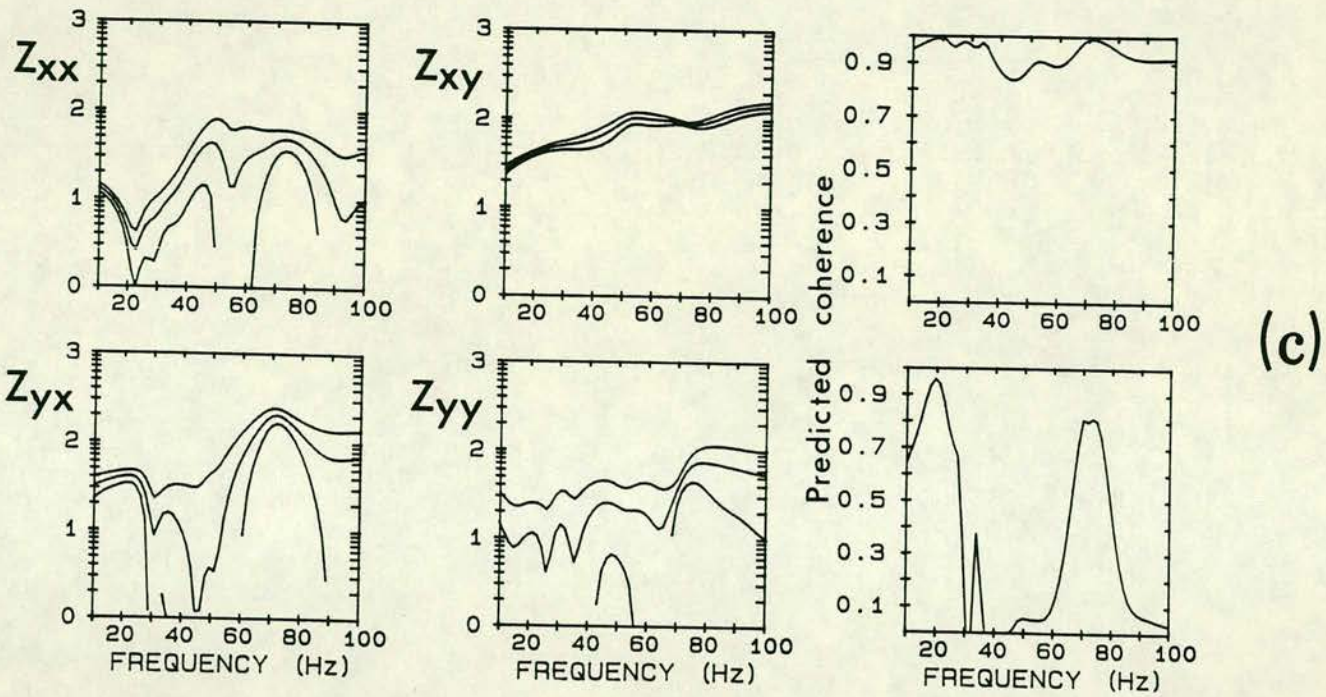


FIGURE 3 Analysis of the Figure 1b window. **(a)** the amplitude response of the whole window (N=150, M=10). **(b)** As per (a) for the 0-0.25sec interval (N=50, M=6). **(c)** As per (b) for the 0.25-0.50sec interval. The parameter M refers to the order of the MAXENT AutoRegressive operator extracted from the data.

transient noise. The main feature here is the smooth drop in coherence for the 0-0.25sec interval (Figure 3b) with the striking exception of a 70Hz signal. The noise source that causes this drop also produces severe bias coupling with a notable and *unsuccessful* attempt of the 70Hz signal to locally rectify it; then, it apparently disappears at least only 0.25sec later, with more coherent and 'normal' results throughout the impedance spectrum. Again it is interesting to study the intriguing scheme of time-local S/N ratio variations that dictate the final time-averaged result.

Equivalent results for the Example site 2 (decade 2) will not presented for brevity. From the time series however, it becomes quite apparent that the only hope for a valid response function from this data window would exist in the 0.25-0.50 sec interval. It turns out that this interval contains high quality, uncontaminated and therefore useful data.

III. A discussion

The above studies indicate that by processing short data lengths, the effects of inadequacies in the data caused by industrial noise surges can be isolated. A second interesting observation that can be made concerns the time-local nature of the '*inconsistent*' noise activity, such as that revealed from the example data 1b and Figure 3(b,c). This portion (0.25-0.75 sec) of the data is apparently not affected by any directly identifiable (e.g. transient) noise source; nonetheless its presence is clear and dramatically displayed in the behaviour of the response functions. Given the extremely time-local nature of the signal

sources as well, (e.g. Beamish and Tzanis 1986, Tzanis and Beamish 1987, in Appendix I), the picture that emerges is revealing. MAXENT appears to be able to extract information and delineate such short term (< 0.25sec) noise contributions.

However, spectral analysis, by itself, cannot recognize inadequacies and large scale inconsistencies and, most important, discriminate them from the data. It must also be appreciated that the computation of response functions from the coherent components of the isolated inadequacies (coherent noise sources), will introduce a population of extreme outliers, (non-Gaussian contributions), that will tend to influence the statistical estimation of the true Earth response. This will be on top of the scatter associated with incomplete or inconsistent data. Unfortunately this will be true for *any* spectral method. The problem becomes more complex because no a priori information is usually available about the 'transmitted' characteristics of transient waveforms, (origin location, source moments, propagation paths), even if their cause is known or understood, e.g. Jones and Kelly, 1966, Schnegg et al, 1986.

Rejection criteria based on spectral power levels are not always warranted, especially in the ELF band, in view of the impulsive and unpredictable nature of the higher signal to noise ratio waveforms. I would like to stress that transient noise waveforms can possess amplitudes comparable to those of some impulsive signals (weak or sufficiently distant sources, e.g. Fig. 1d). Nonetheless, control tests to detect whether the received waveforms satisfy the plane wave assumption can be designed. For instance, the measurement of the radial (atmospheric) electric field together with the horizontal magnetic

components, will theoretically permit the discrimination of distant (>1000Km) from regional or local lightning discharges (Kemp, 1971), and, accordingly, adequate or inadequate waveforms. The Q factors associated with the spectral peaks of ELF observations can be used in the same manner. Q factors of natural processes are of low magnitude (<8 for the Schumann resonances band for instance, Bliokh et al, 1980), reflecting the poor transmissivity of the ionosphere. For lower than the ELF frequencies spatial wavefield information would be required (e.g. Frohlich, 1971). It becomes clear that experimental implementation of this kind of 'control' could be expensive. The weapons to attack such noise problems can first be sought in robust numerical and statistical techniques.

Because of the spatially coherent characteristics of some waveforms, the remote reference method (Gamble et al, 1979a,b) is not always guaranteed to work, let alone that in many cases, truly 'remote' stations may not be available at all. Park and Chave (1984) discussed the use of the Singular Value Decomposition (SVD) approach for the computation of stable, consistent estimates of the impedance function from the observed spectra. Another powerful approach is the robust statistical method of Egbert and Booker (1986). This was presented for GDS transfer function estimation, but can readily be adapted to MT and AMT data processing as well. These procedures may, in general, produce excellent results. The SVD method however is nothing but a numerically stable procedure for solving the LS problem (e.g. Lawson and Hanson, 1974). It will inevitably fail in the presence of large scale inadequate waveforms in the data, since these will tend to dominate the solution eigenvectors. The Egbert and Booker (1986) method would normally be more powerful in such cases. Both approaches are required to solve lengthy

overdetermined systems for single frequency impedance determinations, which is, normally, a quite non-trivial business, and absolutely data dependent. Modern data collection schemes tend to accumulate large numbers of data realisations so that these procedures may be tedious for small computer systems. For those who do not wish to be entangled in the labyrinths of SVD or robust statistics, weighted least-squares (WLS) procedures and weighted averages (WA) as in Stodt (1986), can be considered. These perform well for non-stationary noise sources, or when a small number of outliers are present, but do little to reduce the scatter of the resulting estimates. WLS fails completely when the number of non-Gaussian outliers increases, or when the noise sources force multimodal distributions in the data. Such an effect can easily occur if a persistent noise source exists. The deterioration of weighted LS schemes with increasing non-Gaussian noise content has been described by Egbert and Booker (1986).

All the above thoughts led the author to add his contribution to the complement of anti-noise techniques with the development of two (quite) simple and fast data *selection* procedures. Both techniques work in the time domain. The first comprises a criterion for real-time data quality evaluation. The second can be used prior to frequency transformation, albeit in a post data collection form only. Its primary target is the rejection of outliers and the reduction of scatter, but it can also find a limited application in addressing the problem of multimodal distributions, provided that they stand out well. These data selection techniques are augmented with a study on robust parameter estimation in both the time and frequency domains. An easy-to-use procedure for response function estimation will be produced, such that, even data displaying multiple coherent noise characteristics can yield parameters

with a certain degree of reliability, for single site measurements.

IV. A theoretical basis for the time domain techniques.

Our starting point will be the time domain equivalent of the familiar Magnetotelluric linear system,

$$E_i(\omega) = Z_{ix}(\omega)H_x(\omega) + Z_{iy}(\omega)H_y(\omega) \qquad (1)$$

with i=x,y which, for a finite realisation of the field components of length N, will be given by the discrete convolution

$$e_i(t) = \sum_{\tau=0}^{T} [z_{ix}(\tau)h_x(t-\tau) + z_{iy}(\tau)h_y(t-\tau)] + \varepsilon_i(t) \qquad (2)$$

with i=x,y, assuming the tensor impedance z(t) to be a finite digital matrix wavelet of length T. Lower case letters will henceforth refer to time domain quantities. $\varepsilon_i(t)$ is a noise series that for reasons of simplicity is assumed to reside in the output only. (2) is a form of a 2-input 1-output Wiener LS digital waveform shaping filter equation. The impedance matrix wavelet z(t) is the Earth's impulse response to EM excitation and can therefore be viewed as a filter. Wiener filter solutions proceed with the minimization of $\varepsilon_i(t)$ for all time delays (lags) $\tau = 0,\ldots,T$. The LS solution of (2) minimizes the sum of the squared error (residuals),

$$\frac{1}{N-T} \sum_{t=0}^{N} \varepsilon_i(t)\varepsilon_i(t) = \frac{1}{N-T} \sum_{t=0}^{N} \{e_i(t) - \sum_{\tau=0}^{T} [z_{ix}(\tau)h_x(t-\tau) + z_{iy}(\tau)h_y(t-\tau)]\}^2$$

which in physical terms represents the total noise power contained in the EM field. In the present however we shall only be interested in the $\tau=0$ case. Then, the residual error power in (2) will be given by

$$\frac{1}{N} \sum_{t=0}^{N} \varepsilon_i(t)\varepsilon_i(t) = \frac{1}{N} \sum_{t=0}^{N} \{e_i(t) - \sum_{\tau=0}^{T} [z_{ix}(0)h_x(t) + z_{iy}(0)h_y(t)]\}^2$$

For the LS solution of (2) for z(0) set the partial derivatives of (3)

with respect to $z_{ix}(0)$ and $z_{iy}(0)$ equal to 0, to obtain

$$\text{cov}\{e_i h_x\} = z_{ix}(0)\text{cov}\{h_x h_x\} + z_{iy}(0)\text{cov}\{h_y h_x\}$$

and (4)

$$\text{cov}\{e_i h_y\} = z_{ix}(0)\text{cov}\{h_x h_y\} + z_{iy}(0)\text{cov}\{h_y h_y\}.$$

The operator cov{.} denotes the auto- and cross-covariances of the EM field components, as indicated. For example,

$$\text{cov}\{e_x h_y\} = \sum_{t=0}^{N} e_x(t)h_y(t).$$

For i=x, the equations (4) yield $z_{xx}(0)$ and $z_{xy}(0)$, while for i=y they return the $z_{yx}(0)$ and $z_{yy}(0)$ elements of the z(0) matrix.

Let me now give an interpretation of the z(0) matrix. As stated above the Earth's impulse response is a waveform shaping filter. The impedance waveform shaper has the additional constraint that it must be causal, in the sense that it cannot be determined before the inducing and induced fields exist (i.e. the linear dependence in (2) exists only for preceding time units). The causality requirement imposes minimum delay/phase properties on the impedance wavelet and its frequency response (e.g. Claerbout, 1976) and leads to the minimum dispersion requirement described by Weidelt (1972) and Fischer and Schnegg (1980). Furthermore it can be real since it can relate real quantities. The frequency response (gain) of the impedance wavelet is familiar to EM induction workers as the spectral impedance function $Z(\omega)$, and its phase is the phase of $Z(\omega)$. In a minimum delay system, the energy buildup is concentrated in the first few filter coefficients with the first one usually being the largest in magnitude. By Robinson's (1962) energy delay theorem, the filter should describe the behaviour of a minimally dispersed process required by causality. The z(0) matrix is nothing but the first approximation to the first filter coefficient. It is a *scale factor between the amplitudes of the input and output waveforms, for the linearly related field components at the level of the above*

approximation. Thus, it can be viewed as representing the integrated energy conservation between the magnetic and telluric fields for this 'first order' approximation. The z(0) matrix is calculated through the auto- and cross-covariances of the EM field, and, because the covariance represents *power*, the procedure is similar to the frequency domain LS estimation of the spectral impedance as in Sims et al (1971), where auto- and cross-spectra are used. The two procedures are thus shown to be *isomorphic*.

By considering noise minimisation for the τ=0 case, the filtering problem is reduced to a simple problem of multiple linear regression. This problem can afford some goodness of fit statistics, i.e. the ordinary, partial and multiple correlation coefficients. (Brillinger, 1981, pp 287-295, Kendal and Stuart, 1967, Chapt. 27). Denote the input field components in (2) as $(h_x, h_y) = (h,d)$, and the goodness of fit statistics are given as

(a) multiple correlation coefficient $R^2_{i.hd} = \text{cov}\{\hat{e}_i\hat{e}_i\}/\text{cov}\{e_ie_i\}$, $i=x,y$ where $\text{cov}\{\hat{e}_i\hat{e}_i\}$ and $\text{cov}\{e_ie_i\}$ are the predicted and observed auto-covariances of the output (electric) fields respectively, and

(b) partial correlation coefficients

$$R^2_{ih.d} = (R^2_{i.hd} - R^2_{id}) \,/\, (1 - R^2_{id}) \text{ , and}$$

$$R^2_{id.h} = (R^2_{i.hd} - R^2_{ih}) \,/\, (1 - R^2_{ih}) \text{ ,}$$

with R^2_{ih} and R^2_{id} representing the ordinary correlation coefficients between the input and output channels (e.g. Bendat and Piersol, 1971). Likewise, an expression for the random error associated with the estimates z(0) is afforded as

$$\varepsilon^2_r = \frac{2}{n-2} F_{2,n-2:a} \frac{[\,1 - R^2_{i.hd}\,]\ \text{cov}\{e_ie_i\}}{[\,1 - R^2_{ij}\,]\ \text{cov}\{h_jh_j\}} \text{ , } i = x,y, \quad j = h,d$$

where n is the number of degrees of freedom associated with the measured covariances and $F_{2,n-2:a}$ the 100a percentage point of an F distribution with 2 and n-2 degrees of freedom. One will recognise in the above relations the expressions for the partial and multiple coherences of frequency domain multiple-input/single output linear systems (Bendat and Piersol, 1971) as adapted for MT data processing (Reddy and Rankin, 1974), and the associated random errors (Goodman, 1965), with the spectral estimates substituted for covariances. The aforementioned isomorphism is sufficient to explain the similarities and relieve the author from the obligation to derive them; the methods are real versions of the ones used by the quoted authors, and the statistics identical. A demonstration of the derivation of the random errors for the simple case of the 1-input/1-output problem, is provided in the Appendix 4.A to this chapter.

Of the goodness of fit statistics presented above, the multiple (predicted) correlation coefficient is of particular significance and can be derived while solving for z(0). Expand the squared term in the right hand side of (3) and perform the summations to obtain

$$\begin{aligned} \mathrm{cov}\{e_i e_i\} - \mathrm{cov}\{\varepsilon_i \varepsilon_i\} = \; & 2[z_{ix}(0)\mathrm{cov}\{e_i h_x\} + z_{iy}(0)\mathrm{cov}\{e_i h_y\}] \\ & - 2\, z_{ix}(0) z_{iy}(0)\mathrm{cov}\{h_x h_y\} \\ & - z_{ix}^2(0)\mathrm{cov}\{h_x h_x\} - z_{iy}^2(0)\mathrm{cov}\{h_y h_y\} \end{aligned} \qquad (5)$$

The left hand side of (5) is simply $\mathrm{cov}\{\hat{e}_i \hat{e}_i\}$. In the next section, I will discuss the use of the predicted correlation coefficient, for a rapid synoptic evaluation of data quality, as well as the utilization of the z(0) factors for data selection. In the following, the convention $z_{ij} = [z(0)]_{ij}$,i,j=x,y will be used, to denote the z(0) matrix elements.

IV.1 Inclusion of a reference channel.

The above analysis can easily be extended to cover additional EM field recordings, for remote reference applications. Our starting point will again be equation (2). Denote the reference channel by r(t). Then multiply both sides of (2) with r(t) to obtain

$$e_i(t)r(t) = \sum_{\tau=0}^{T} [z_{ix}(\tau)h_x(t-\tau) + z_{iy}(\tau)h_y(t-\tau)]r(t) + \varepsilon_i(t)r(t)$$

Let T=0, then take the cross covariances of both sides over the time interval N, to obtain

$$\frac{1}{N}\sum_{t=0}^{N} e_i(t)r(t) = \frac{1}{N}\sum_{t=0}^{N}[z_{ix}(0)h_x(t) + z_{iy}(0)h_y(t)]r(t) + \sum_{t=0}^{N}\varepsilon_i(t)r(t)$$

which gives

$$\text{cov}\{e_i r\} = z^r_{ix}(0)\text{cov}\{h_x r\} + z^r_{iy}(0)\text{cov}\{h_y r\} + \text{cov}\{\varepsilon_i r\} \qquad (6)$$

However, under the usual assumption that the noise term $\varepsilon_i(t)$ and the reference channel are uncorrelated, we have $\text{cov}\{\varepsilon_i r\}=0$, and (6) reduces to the form of (4), with no auto-covariances included. The multiple correlation coefficient will be found as with the single site case, by inserting the $z^r(0)$ determinations in (5), instead of the local z(0). The variance of $z^r(0)$ can be obtained with the real version of the analysis presented by Gamble et al (1979b). Because I am not going to present any applications with reference channels, I shall not investigate the matter any further. Note however, that the details of the following analysis and investigations, are directly applicable to this remote reference extension of the basic approximation.

V. Two time domain data selection techniques.

V.1 The use of the multiple correlation coefficient for data selection (Technique A).

Modern AMT data processing, based on digital computers, manipulates large numbers of usually automatically recorded data realisations. The state of the art procedure for rejecting data windows degraded by noise, is to obtain the predicted coherence spectrum, and test against a normally preset threshold. This procedure requires a minimum of four frequency transformations per data window (one per data channel), usually computed with the FFT procedure, and, may be computationally expensive, especially in high frequency, high speed surveys, where time is a critical parameter and small computers and on-line processing techniques are employed. The fast assessment of data quality becomes important because of the time-local behaviour of some noise sources, as described above, accentuated in the high (>100Hz) AMT frequency range. The time required to transform and test the incoming data, reduces the probability of exposure to more useful information, necessary to stabilize the required impedance functions. Time limit requirements may also be present during off-line processing of broadband multi-station recordings and their associated large quantities of data.

In such cases, a rapid evaluation of data quality can be afforded by the multiple (predicted) correlation coefficient R_i^2 ,i=x,y. The computational expense of obtaining R_i^2 has been tested against the same for obtaining the predicted coherence function $\gamma_i^2(\omega)$,i=x,y. The FFT algorithm used for frequency transformation is the NLOGN routine of Robinson (1967). Comparisons of CPU time on a PDP-11/23 computer revealed a speed ratio of 13:1 for the time domain test.

The utility of the R^2 test can be evaluated through a comparison with

the conventional predicted coherence test. It must be appreciated that there can be no unique relationship between these quantities, since R^2 is a scalar quantity and $\gamma^2(\omega)$ is a function of frequency. A degree of correspondence can be afforded however, if we use the arithmetic mean of $\gamma^2(\omega)$ over the bandwidth of analysis, i.e.

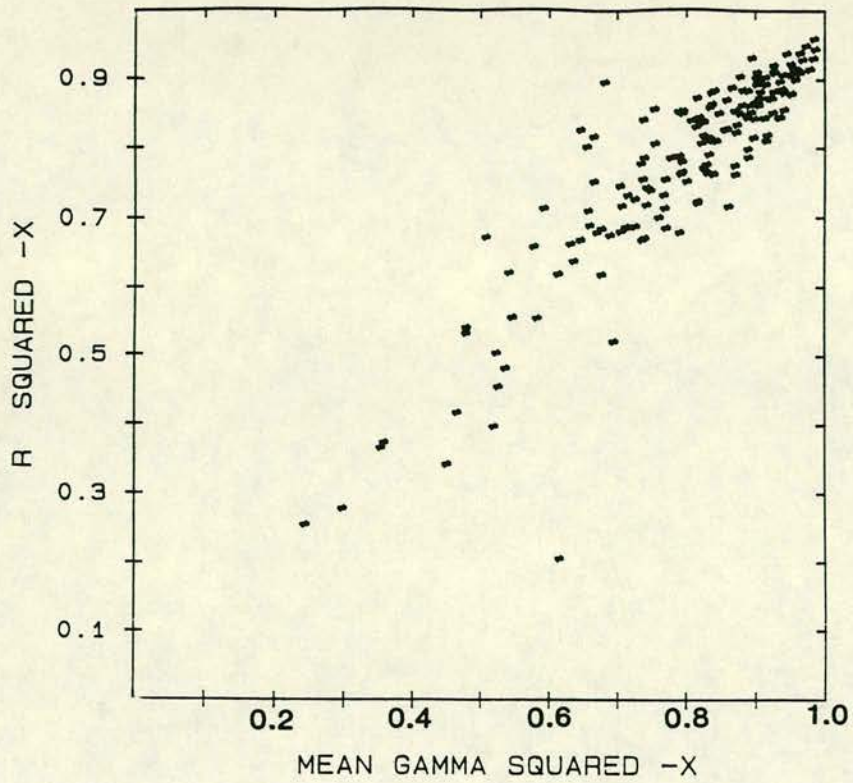
$$\langle\gamma_i^2\rangle = \left(\sum_{k=1}^{N\omega} \gamma_i^2(\omega)\right) / N_\omega$$

where N_ω is the number of available frequency estimates. It is important to note here that for complete waveform shaping, i.e. for complete estimation of the $\mathbf{z}(t)$ matrix wavelet, the relationship $R_i^2 = \langle\gamma_i^2\rangle$, i=x,y holds. Since, in our case, R_i^2 is the correlation coefficient for the first approximation, this equality will not be true, except for the case of induction over a uniform conducting half space. Figure 4 displays the results of plotting R_i^2 vs $\langle\gamma_i^2\rangle$ for a set of 165 data windows of decade 1 data. Figure 4a displays the result for i=x, and Figure 4b the same for i=y. It is immediately apparent that a quasi-linear relationship exists between the two quantities. The values of R_i^2 are generally lower than their corresponding $\langle\gamma_i^2\rangle$ values. This is partly due to the absence of a unique relationship between the two quantities. The latter can also account for the scatter observed, especially in the lower end of the available distribution. However, this behaviour of R_i^2 can be expected because it is derived through total energy conservation considerations. It will therefore be very sensitive to non-linear features in the data. These will take the form of perturbations of the covariance functions and can produce very low values for R_i^2.

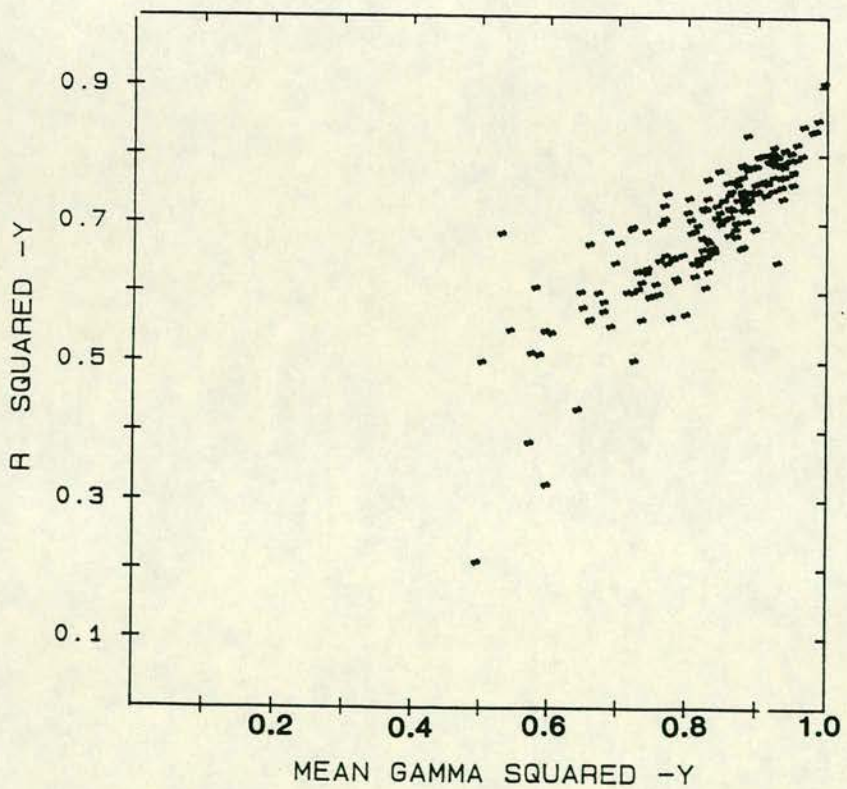
In order to understand the behaviour of R^2 consider the data window of Figure 5a, displaying decade 1 data. The corresponding frequency domain LS solutions (Sims et al, 1971) are produced in Figure 5b for the

downward biased off diagonal elements. This data window displays characteristics typical of the transient Schumann resonances sferic waveforms (e.g. see Tzanis and Beamish 1987, in Appendix I). The initial part of the large amplitude waveform represents the arrival of the direct (short path) signal from a powerful lightning discharge. The subsequent fluctuations are due to the arrival of the long path signal and to the circulation of energy arround the globe. In this data, the e_x short path signal comprises a curtailed sferic in the form of an incomplete quasi-sinusoid, while the short path arrivals in the other channels are almost perfect damped quasi-sinusoids. The long path signals are naturally of smaller amplitude, more dispersed and apparently contain more noise. As can be seen, the predicted coherences for i=x are high (>0.85) for the 20Hz-60Hz bandwidth and quite low otherwise. This is due to the quality of both the long and short path signals in the e_x telluric channel. The breakdown of coherence accumulates during integration, when the covariances are calculated, and is reflected in the low value of R_x^2 (0.25) returned. On the other hand, the i=y (Z_{yx}) solution displays a high degree of coherence throughout the recorded bandwidth, with a relative reduction towards the high frequencies. This is produced mainly by the noise terms in the long path signal and the rest of the low amplitude waveforms. When accumulated this causes the R_y^2 to drop to 0.75. The degradation of this data by noise is best observed through the distorted covariances, while the predicted coherence values are kept high because of the better quality and intensity of the short path signal. The erratic nature of the impedance spectrum, especially for the Z_{xy} element, appears to support this notion.

It is evident that R^2 is a very sensitive measure of data quality.

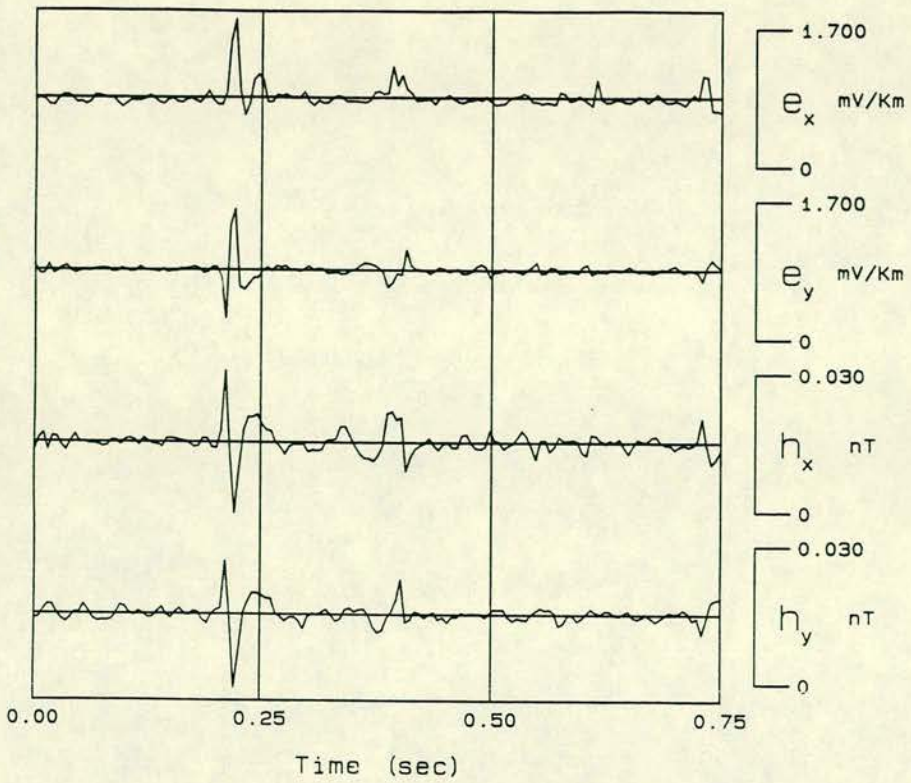


(a)

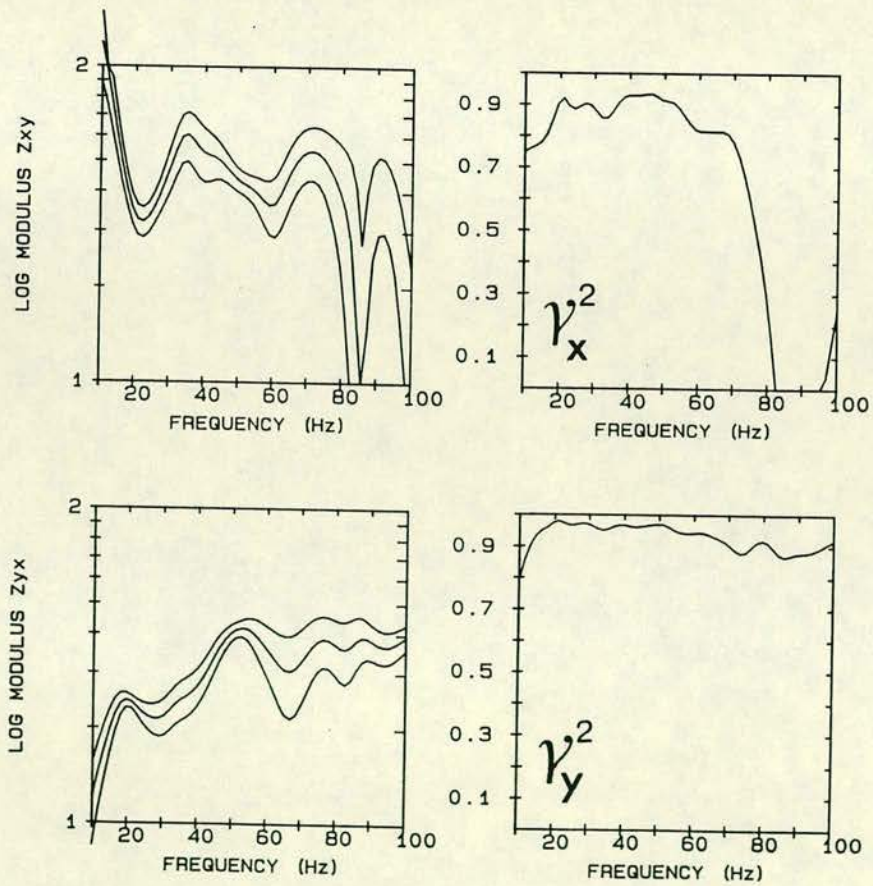


(b)

FIGURE 4 Two plots of R_i^2 vs $\langle\gamma_i^2\rangle$, i=x,y for a set of 165 decade 1 data windows. **(a)** displays results for the i=x solutions; **(b)** displays results for the i=y solutions.



(a)



$R_x^2 = 0.25$

(b)

$R_y^2 = 0.75$

FIGURE 5 **(a)** One window of decade 1 data. Sampling interval is 200Hz; effective bandwidth is 100Hz-10Hz. **(b)** The corresponding frequency domain impedance determinations for the off-diagonal downward biased elements and their associated predicted coherence functions. Maximum entropy spectra (N=150, M=10). Linear frequecy scale.

Threshold values of R^2 >0.60 will in general return predicted coherence values in excess of 0.85, while R^2 >0.40 normaly guarantees predicted coherences over 0.65. The rapid degeneration of R^2 with increasing noise content is due to the fact that it reflects the linearity conditions for the first approximation only. If one proceeds to obtain complete waveform shaping (higher order wavelets), the goodness of fit will improve, but more complicated programing will be required, and, speed and simplicity will be lost. The sensitivity and simplicity is exactly the power of the criterion. The first order goodness of fit is important in that it provides a first clue as to what is to be expected from a particular data window. Therefore, while high R^2 thresholds will accept excellent data, lower values can be used, depending of course on the noise regime in the vicinity of the recording station. The author has used thresholds of the order of 0.20 - 0.30 for data suffering from intense noise interference. For normal sounding conditions, an acceptance scheme such as R_i^2 >0.40, i=x,y will in general return very satisfactory results. Higher thresholds are not recommented in general.

Note that the proposed criterion does not aim to replace the conventional predicted coherence tests altogether, but to supplement it. Analysis will still be done in the frequency domain, at least for the foreseeable future. It is a fast and simple technique that can be used to test the incoming data, so that frequency domain analysis will be applied to those realisations, guaranteed to contain at least some 'acceptable' results.

V.2 The use of the z(0) factors for data selection (Technique B).

Let us begin with the usual assumption that the random noise content in

all recorded channels is statistically independent. It is therefore conceivable that the elements z_{ij} of the z(0) matrix will belong to statistically independent populations for a number of observations L. The distribution of the z_{ij} element population will depend on the nature of the noise contained in the observations. For noise free data all the determinations of z_{ij} should be time invariant and absolutely consistent. For data contaminated with random (Gaussian) noise, the z_{ij} populations can be expected to obey a normal distribution law for large L, with the spread of the the Gaussian bell increasing with increasing noise content; noise introduces scatter in the observations. The independent z_{ij} populations will produce their own statistics, so that the mean (<z>) and standard deviation (s) matrices can be formed as

$$\langle \mathbf{z} \rangle = \begin{bmatrix} \langle z_{xx} \rangle & \langle z_{xy} \rangle \\ \langle z_{yx} \rangle & \langle z_{yy} \rangle \end{bmatrix} \qquad \mathbf{s} = \begin{bmatrix} s_{xx} & s_{xy} \\ s_{yx} & s_{yy} \end{bmatrix}$$

The assumption of normally distributed errors also permits the calculation of confidence intervals for <z>, based on a t-distribution statistic.

The introduction of inadequate coherent waveforms introduces a new parameter for consideration, i.e. their statistical behaviour. Obviously, these waveforms will produce solutions that do not correspond to local passive induction processes, and their 'impedances' will not be consistent estimates of the true Earth response functions. Instead, they will tend to stand out as outliers of the (adequate) impedance populations. This will apparently be so for the z_{ij} factors as well. If the inadequate noise sources are random, outlier populations will tend to form arround the tails of the true z_{ij} Gaussian bells; the more the sources the more heavily tailed these distributions will be. If the inadequate noise sources repeat, we may face the formation of multimodal

distributions in the observed populations. In view of its nature, the z(0) matrix has the advantage of offering a synoptic evaluation of the consistency of the data set. Thus, the z_{ij} populations can be expected to provide a summary of the data and noise information contained in the observations.

This information can be displayed in a number of ways. Figure 6 depicts the time history of the z_{ij} factors from Example Site 1, together with the predicted (multiple) correlation coefficient. The 'time history' is just the series of the z_{ij} factors, computed sequentially from the incoming data windows (observational series). One impressive, and I believe quite surprising first observation, is that the series are structured, and can even be subjected to a basic harmonic analysis. The thick continuous curves indicate a polynomial of the 5'th degree, that can be easily extracted from the observational series. Further scrutiny reveals that the most consistent 'long period' structural feature, approximately between observations 41 and 90, is due to a clustering of noise waveforms, including transient impulsive events with variable amplitudes, that arrive within this time period. Sporadic transient waveforms can be seen as outliers (spikes) over all the observational series, and in particular in the z_{yx} and z_{yy} elements. These display less prominent 'long period' structure and appear to be basically infested by the powerful transient waveforms. The last part of the series (between observations 123 and 164) are dominated by a sequence of noisy spikes. Due to the existence of this cluster, the 'long period' structure depicted in the polynomial trend should not be taken at face value. The z_{xx} and z_{xy} elements are considerably more structured; the repetitive form of their observational series makes one wonder whether this can be attributed to a periodic nature of the anthropogenic activity that

produces the noise waveforms. The E-W and N-S measuring directions in this site, apparently suffer from quite different, directional noise regimes. More insights can be provided through a simultaneous study of the R^2 series; this is, in general, quite high (>0.40) for the z_{xx} and z_{xy} elements, which, according to V.1, would guarantee predicted coherences in excess of 0.70 for most of the recorded data windows. It is also quite apparent that R^2 follows the trends of the z_{ij} series. Thus, the R_x^2 values are consistently higher for the noise population clusters, particularly in the 41-90 segment of recorded data windows. Also, note that several high value excursions of R_x^2 coincide with the outliers of the z_{xx} and z_{xy}series. This feature is much more apparent in the z_{yx} and z_{yy} elements. Here, we have quite low correlation coefficients, that suddenly jump to high values when the noise transients arrive, with only a few exceptions. The exceptions can actually be traced down to 'normal' noisy data. This clearly demonstrates the coherent characteristics of the noise waveforms. If the conventional estimation procedures are applied to the data of this site, i.e. selection on the basis of predicted coherence tests only, then one would end up enhancing the noise content of the data, by accepting more noise waveforms, with all the deleterious consequences implied in the analysis and interpretation of the resulting impedance functions.

The distributions of the z_{ij} factors, can also be presented in the form of histograms, such as the ones shown in Figure 7(a-d), for the Example Site 1; they contain determinations from 165 data windows with N=150. 'Noise' information is conveyed through the spread of the magnitude of the z_{ij} determinations as a function of their frequency of occurrence within a class interval set equal to half a standard deviation. This is

a very effective general way of presenting populations about which no prior information exists. With a finer class interval, more features of the data structure could be observed, however, this presentation provides sufficient information for a basic insight into the data. Firstly, it can be seen that for the given approximation, the z_{ij} populations do not, as expected, form normal distributions. Noisy populations tail off to the right and left of the major clusters of the z_{xx} and z_{xy} populations respectively. The powerful directional transients manifest themselves to the left of the z_{yx} major cluster. Moreover, the spread of this element population is quite narrow indicating an otherwise normal noise environment. In all cases, the observed $\langle z_{ij} \rangle$ and that of the major clusters of the z_{ij} populations are not coincident.

Given the basic z(0) element statistics ($\langle z \rangle$, s), and the visual information provided by the histograms and the observational series, I propose a data selection procedure in which a data window shall be accepted if

$$R_i^2 > \text{threshold}, \quad i=x,y$$

and

$$\text{(A)} \quad (\langle z_{ij} \rangle - \alpha s_{ij}) \leq z_{ij}(l) \leq (\langle z_{ij} \rangle + \alpha s_{ij}), \qquad l=1,\ldots,L,$$

or

$$\text{(B)} \quad \text{left}(z_{ij}) \leq z_{ij}(l) \leq \text{right}(z_{ij}), \qquad l=1,\ldots,L,$$

Both schemes will accept data subject to 6 independent constraints. A data window is accepted only if all four of its z(0) solutions lie within the optimum ranges defined by their individual population statistics (scheme A, with α a real constant), or user provided bounds (scheme B), and, a threshold for the multiple correlation coefficient is satisfied. Therefore, in a computer program only scheme (A) can be

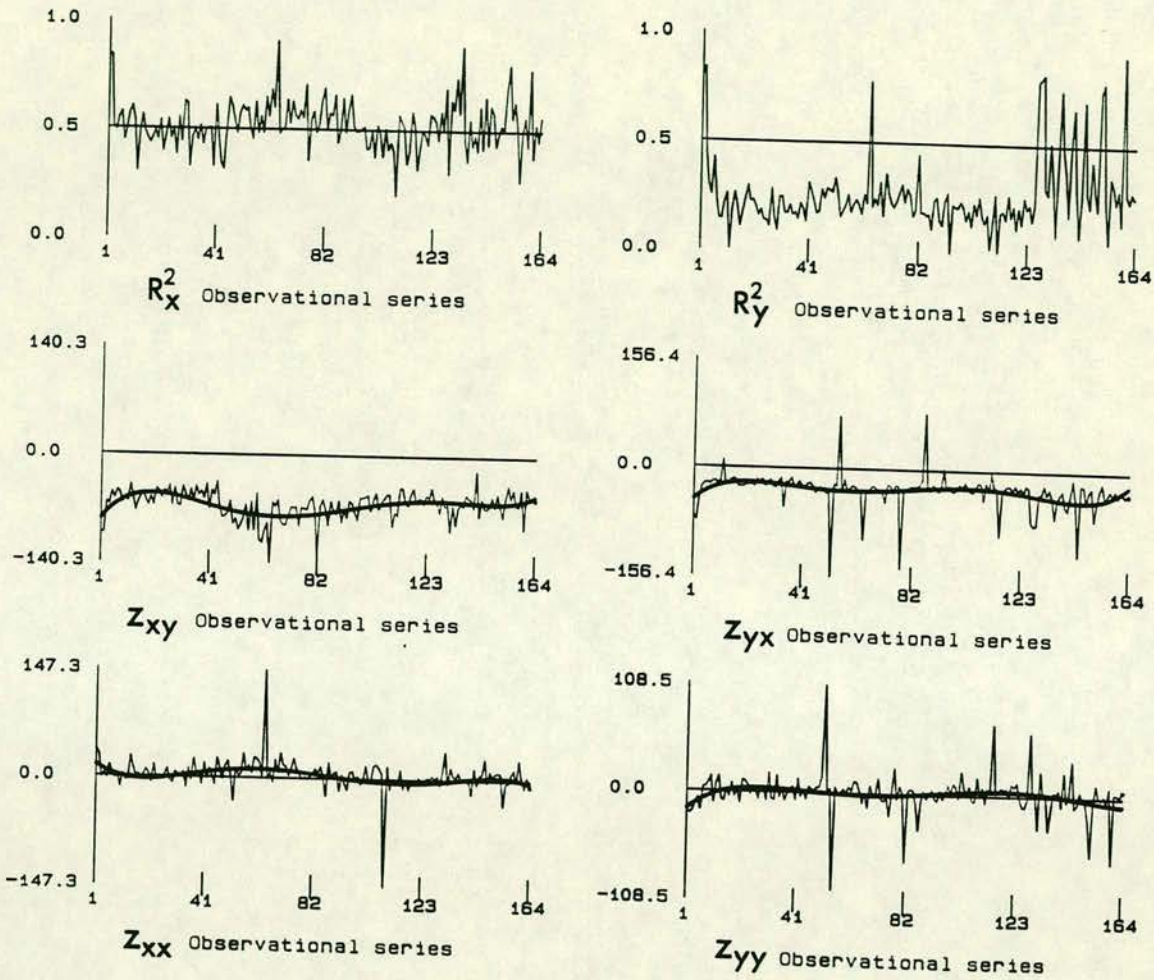


FIGURE 6 The z_{ij} factors, and the multiple correlation coefficient for Example Site 1, arranged in sequential form (observational series). The thicker solid curves indicate the underlying structure (trends) of the series and correspond to polynomials of degree 5.

(a)

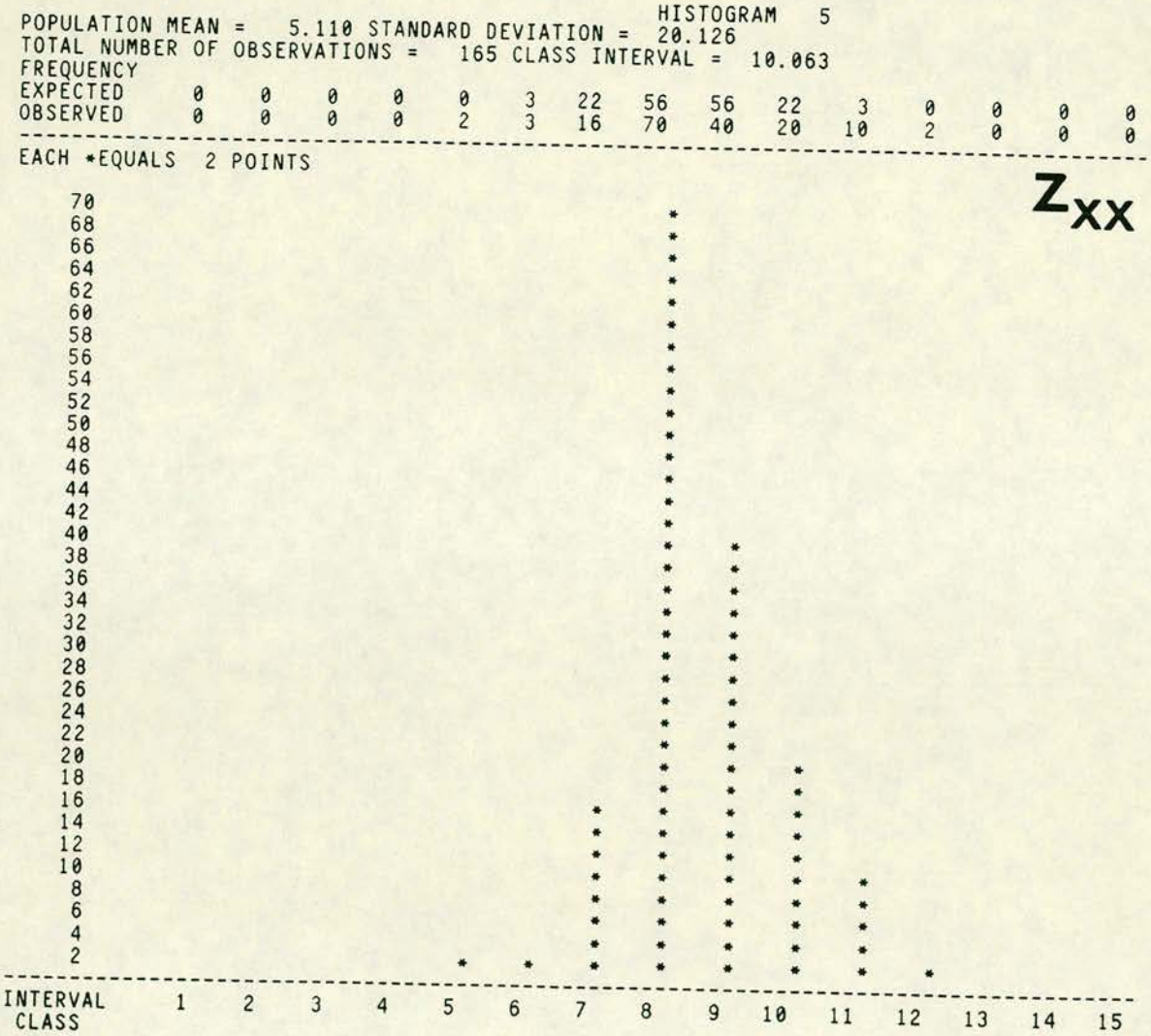


FIGURE 7 **(a)-(d),** the z_{ij} factors displayed in histogram form (Example Site 1). 'Expected' frequency of occurence is that of a normal distribution.

(b)

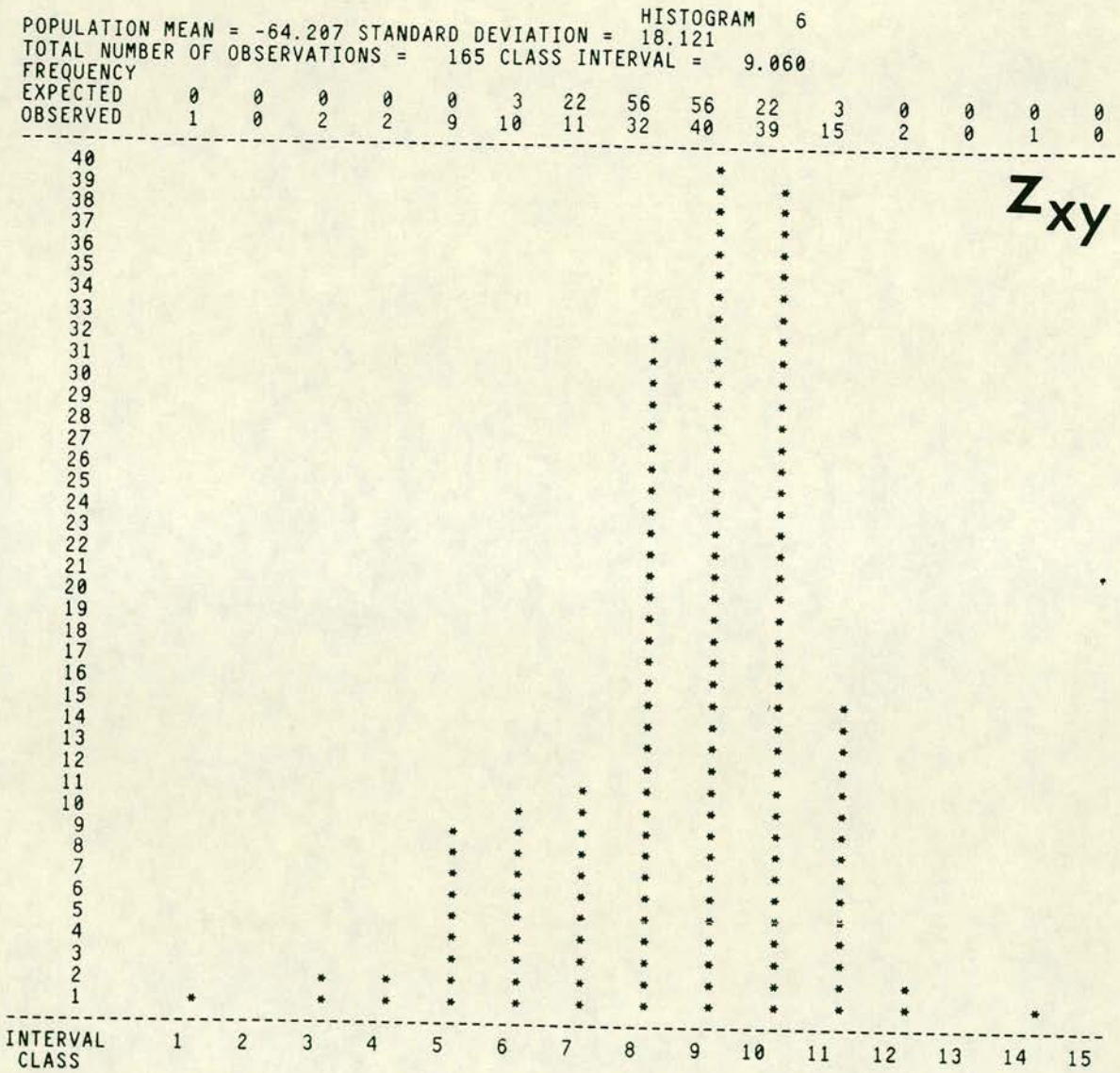


FIGURE 7 ..continued

(c)

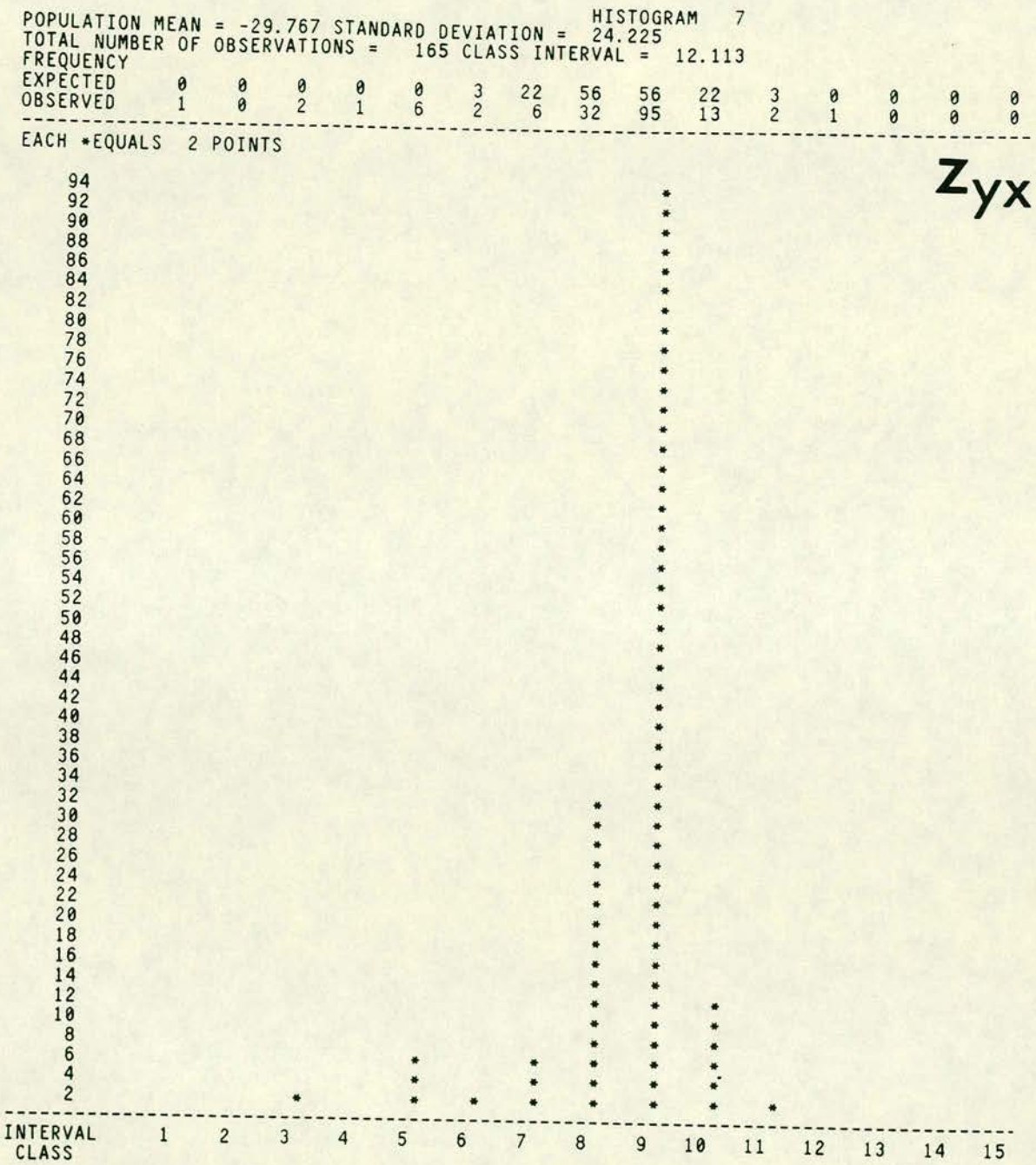


FIGURE 7 ..continued

(d)

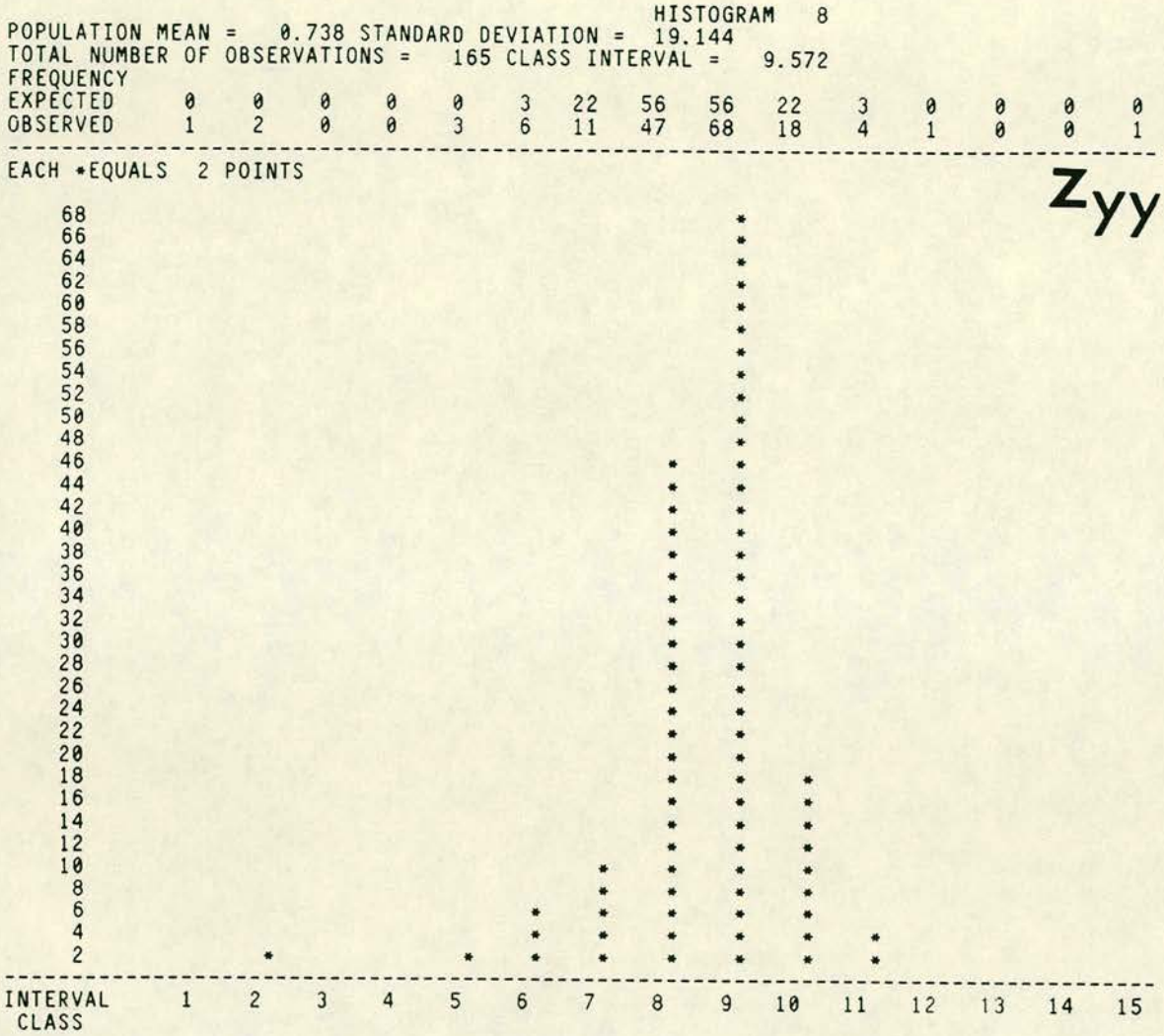


FIGURE 7 ..continued

fully automatic, while scheme (B) requires a prior inspection of the histograms.

VI. Parameter estimation in intense, or non-Gaussian noise environments.

Because of the isomorphism between the first order time and frequency domain processes, all the data properties and estimation procedures to be discussed are applicable in *both*. Reminders of this fact will be quite frequent. However, because of the inherent simplicity of the time domain process, (only real numbers are involved), all investigations will be carried out in this domain. The minor differences in computational requirements, between the two domains, will be discussed in due course.

The crucial element in the implementation of the automatic scheme (A) is the calculation of reliable estimates for the $\langle \mathbf{z} \rangle$ and $\mathbf{s}$ matrices. Equations (4) define two constraints for the LS solutions for each pair of the z_{ix}, z_{iy}, i=x,y elements of the z(0) matrix. For a number of data windows L, there exist 2L such constraints (observational equations) that comprise the overdetermined system

$$\mathbf{a} = \mathbf{b}\,\mathbf{z} + \underline{\varepsilon} \qquad (7)$$

or in an explicit form,

$$\begin{bmatrix} \mathrm{cov}\{e_i h_x\}_1 \\ \mathrm{cov}\{e_i h_y\}_1 \\ \mathrm{cov}\{e_i h_x\}_2 \\ \mathrm{cov}\{e_i h_y\}_2 \\ \cdots \\ \mathrm{cov}\{e_i h_x\}_L \\ \mathrm{cov}\{e_i h_y\}_L \end{bmatrix} = \begin{bmatrix} \mathrm{cov}\{h_x h_x\}_1 & \mathrm{cov}\{h_y h_x\}_1 \\ \mathrm{cov}\{h_x h_y\}_1 & \mathrm{cov}\{h_y h_y\}_1 \\ \mathrm{cov}\{h_x h_x\}_2 & \mathrm{cov}\{h_y h_x\}_2 \\ \mathrm{cov}\{h_x h_y\}_2 & \mathrm{cov}\{h_y h_y\}_2 \\ \cdots & \cdots \\ \mathrm{cov}\{h_x h_x\}_L & \mathrm{cov}\{h_y h_x\}_L \\ \mathrm{cov}\{h_x h_y\}_L & \mathrm{cov}\{h_y h_y\}_L \end{bmatrix} \begin{bmatrix} z_{ix} \\ z_{iy} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \cdots \\ \varepsilon_{2L-1} \\ \varepsilon_{2L} \end{bmatrix}$$

with i=x,y. In general, one seeks to solve (7), and obtain a maximum likelihood estimate of $\mathbf{z}$, by minimizing an expression of the form

$$\sum_n \rho(r) = \sum_n \{[a(n) - \mathbf{b}(n)z]/\sigma\}, \qquad n = 1,2,....,2L$$

where $\rho(r)$ is a loss function suitable for the problem in question, r refers to the residual of an observation, and σ is an error scale factor to normalize the residuals. An expression of the form $\rho(r) = |r|$, leads to minimization of the summed absolute deviation (L_1 norm); this type of solution will not be examined here. Standard LS techniques use $\rho(r) = r^2/2$, i.e. minimize the sum of squared residuals (L_2 norm)

$$\sum_n r^2(n) = \sum_n [a(n) - \mathbf{b}(n)z]^2$$

which leads to the formation of the 2x2 system of normal equations

$$(\mathbf{b}^T\mathbf{a}) = (\mathbf{b}^T\mathbf{b})\, z$$

with the solution $z = (\mathbf{b}^T\mathbf{b})^{-1}(\mathbf{b}^T\mathbf{a})$. If it so happens that there exists some kind of measure of reliability of the observations, one may desire to weight the more reliable equations so that they influence the calculations more heavily. This corresponds to the minimization of the weighted sum of the squared residuals

$$\sum_n w(n)r^2(n) = \sum_n w(n)[a(n) - \mathbf{b}(n)z]^2 ,$$

The best linear estimator for z will be given by

$$z = (\mathbf{b}^T\mathbf{W}\mathbf{b})^{-1}\,(\mathbf{b}^T\mathbf{W}\mathbf{a})$$

with $\mathbf{W} = \mathbf{w}\mathbf{w}^T$, $\mathbf{w}^T = [w(1)\ w(2)...w(2L)]$, and the assumption $w(i)w(j)=\delta_{ij}$ is made. This constitutes the Weighted Least Squares (WLS) algorithm. The classical method to solve (7) is to solve the system of normal equations , which implies the inversion of $(\mathbf{b}^T\mathbf{b})$. However, there exist other, more accurate ways for the solution of the LS problem. A survey of commonly available literature (e.g. Claerbout 1976, Golub and van Loan 1983, Press et al 1986, and others) indicates that at least twice as much precision is required in order to invert $(\mathbf{b}^T\mathbf{b})$, than to deal directly with **b**. Golub (1965) offered an alternative approach to the solution of (7), by reducing **b** to an upper triangular form with

Householder transformations, and thereafter obtaining the solution vector with backsubstitutions. In this analysis, both methods will be considered. The computer program for the formation and solution of the normal equations is due to Malin et al (1981), while the code for Golub's method has been taken from Claerbout (1976). The application of both methods to the data of the Example Site 1 returns identical results, i.e.

$$z_{xx} = 35.45 \qquad z_{xy} = -123.74$$

$$z_{yx} = -72.78 \qquad z_{yy} = 7.30$$

which are rather hilarious. This total failure is due to the extreme outliers evident in the observational series of Figure 6, that dominate the formation of the covariance matrices (normal equations), and the internal products during the Householder transformations. The magnitude of the failure can easily be seen in the histograms of Figure 7, where the population arithmetic means are given as

$$z_{xx} = 5.11 \pm 20.18 \qquad z_{xy} = -64.20 \pm 18.17$$

$$z_{yx} = -29.76 \pm 24.76 \qquad z_{yy} = 0.78 \pm 19.20$$

A more detailed study of the histograms actually shows that the arithmetic means are also biased quantities. The most clear example is that of the z_{xy} element, which is infested by the population tailing off to the left of the main cluster of estimates (Figure 7b). It is clear that the true mean of the adequate data should be expected somewhere inside class 9. In fact, a study carried out with more detailed versions of the histograms of Figure 7 shows that the true mean should be expected towards the boundary of classes 9 and 10. This is clearly quite far from the arithmetic mean as well. Similar effects infest the rest of the z_{ij} determinations.

The direct application of standard LS prodedures is therefore out of the

question. This method will work perfectly well with data possessing normally distributed errors, but will inevitably fail when powerful extreme outliers are present, and/or consistent alien populations. In the case of Example Site 1 we have both. The situation is agravated by the fact that no unequivocal measure of reliability exists, in order to weight the observational equations; the predicted correlation coefficient (coherence) is certainly not the appropriate quantity.

An alternative approach is to use the individual z_{ij} populations, to derive weighted averages (WA) by expressions of the form

$$\langle z_{ij} \rangle = \sum_{l} w_{ij}(l)\, z_{ij}(l) \left[\sum_{l} w_{ij}(l) \right]^{-1} \qquad (8)$$

where $w_{ij}(l) = 1/\varepsilon^2_{ij}(l)$, i=x,y, and l = 1,2,...,L. The bracketed quantity refers to the ensemble average, while ε^2_{ij} is the random error. A measure of the scatter of the data population can be provided by the standard statistics (variance, standard deviation). Implementation of this particular WA procedure yields

$$z_{xx} = 3.31 \pm 20.18 \qquad z_{xy} = -62.72 \pm 18.17$$

$$z_{yx} = -29.49 \pm 24.76 \qquad z_{yy} = 0.71 \pm 19.20$$

which is somewhat closer to the expectation values, for the z_{xx} and z_{xy} solutions, but not spectacularly better either. The efficiency of the random error (variance) as weight factors however, deserves some more consideration. Observe that their magnitude largely depends on the power content of the E- and H- fields, and primarily on $\mathrm{cov}\{e_i e_i\}$ i=x,y, i.e. the auto-covariance of the E-field. The value of this quantity can be large and erratic for inadequate data, such as the cases presented herein. The time history of the random errors associated with the z_{ij} factors of Example Site 1, are presented in Figure 8; the predicted correlation coefficient series are also included for reference purposes. This figure demonstrates what is expected, i.e. larger

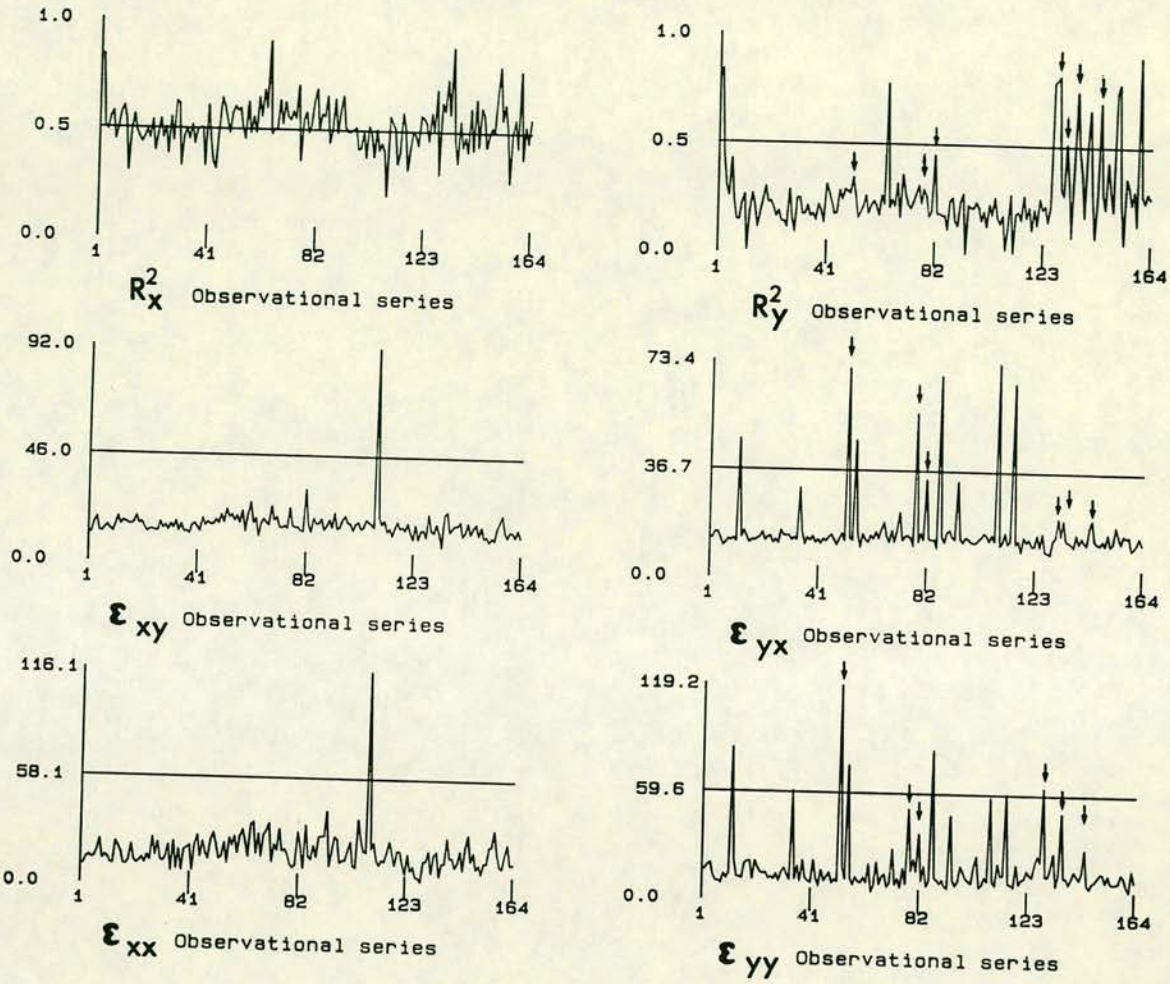


FIGURE 8 The random errors ε_{ij} of the z_{ij} factors, and the multiple correlation coefficient, arranged sequentially as per Figure 6 (Example Site 1).

errors for the less correlated data, and smaller errors for the more correlated. It also shows what may be rather unexpected, i.e. quite large errors for noisy data, correlated with respect to the general properties of the sample space (pointed by arrows). Naturally, because of the virtually unlimited possibilities of noise waveform realizations, no rules are possible; only chances. Therefore, not all correlated noise waveforms produce large random error, e.g. the noise cluster between observations 123-165, for the z_{yx} and z_{yy} factors. This could be attributed to the existence of powerful noise waveforms in both telluric and magnetic channels (e.g. Figure 1b), i.e. not confined to the telluric channels only (e.g. Figure 1c). The small random errors result in the insignificant improvement in the estimation of these factors, upon using the WA estimator. However, since correlation (coherence) is definitely not our best bet for a data quality indicator, random error weighting helps to reduce the effects of some extreme outliers. The WA estimator is a more reliable procedure, as can be seen in the examples given above. A more complete study of the merits of random error (variance) weighting, for very coherent noise waveforms (frequency domain analysis, recall the isomorphism) is given in Appendix 4.B. Coherence is not an unequivocal indicator of data quality, as has been appreciated before (e.g. Kröger et al 1983, and others), but not much has been done about it. This study, therefore, will attempt to produce a procedure to resolve the question of multiple (or even remote) coherent noise induced bias, in both the time and frequency domains, and will do so by heavily relying on robust statistics, and the use of the random error.

VI.1 Robustification of the estimation procedure.

The WA procedure offers a better approximation to the expectation values of the adequate data populations, but it is not a panacea, and, it can be seen that it cannot provide an estimate of the adequate z_{ij} population scatter. Another, and quite important requirement is that one would like to implement an estimation algorithm capable of returning reliable parameters, so that subjective judgement would be minimized. Robustification of the estimation procedure is therefore necessary; the following few pages are an exercise of robust parameter estimation in cases of heavy noise contamination.

We are concerned with the problem of obtaining reliable estimates of the expectation values of the z_{ij} factors, together with reliable estimates of the data scatter. Thus, the estimation procedure must show minimal bias (B-robust) and be associated with a minimal change of variance (V-robust) once the optimal $\langle z_{ij} \rangle$ is achieved. Thus we can define the problem as being one of estimating the location (L) of the parameters that maximize the likelihood function of a set of observations {X}, i.e. we address the problem of defining a maximum likelihood estimator (M-estimator) of parameter location, that is also V-robust. The approach to M-estimation usually requires the solution of a non-linear system of equations. However, this is avoided by implementing iterative procedures that influence and modify the observations, until convergence to the optimal parameters is achieved. Modification of the observations is, in general, performed with the so called influence function $\psi(r)$, which is generically related to the form of the loss function considered, and/or weight factors of the type $w[\psi(r)]$. Theoretical details can be found in Huber (1981) and more recently, in the influence function approach of Hampel et al (1986). The general form of the iterative procedure is as follows:

(1) Obtain an initial estimate for the solution vector z, $\hat{\mathbf{z}}(0)$.

Then, for k=0,.....

(2) Compute the predicted output observation

$$\hat{a}(n,k) = \mathbf{b}(n,k)\ \hat{\mathbf{z}}(k)$$

and the residual

$$r(n,k) = a(n,k) - \hat{a}(n,k),$$

or the standardized residual

$$t(1,k) = r(n,k)/\ \sigma(k),$$

with $\sigma(k)$ usually being the RMS error of the residuals.

(3) Compute the influence function $\psi[r(n,k)]$, and/or the weight function $w(\psi)$.

(4) Modify the output observation as

(i) $a(n,k+1) = \hat{a}(n,k) + w(\psi)\ r(n,k)$, or

(ii) $a(n,k+1) = \hat{a}(n,k) + \psi[r(n,k)]\ t(n,k)$

(5) Obtain new solutions $\hat{\mathbf{z}}(k+1)$ based on the modified observations, and a new error scale $\sigma(k+1)$.

Then, repeat from (2) until convergence. The successive approximations z(k) to the expectation values of z are usually obtained with LS algorithms. This constitutes the procedure of regression-M estimation detailed in Huber (1981), Hampel et al (1986) and others. As will be seen shortly, this is a variant of the WLS procedure. In the present analysis, I will consider the robust solution of (7), based on LS algorithms, as well as WA estimators of location (W-estimators, e.g. Hampel el al, 1986).

The influence (and weight) function, whatever its form, must always approximate r for small r, resemble a damped version of r for larger r, and totally reject gross outliers. The basic choice is between weight-functions that vanish outside some central region, or, are

continuously decreasing as a function of r. The former choice are, in general, the redescending ψ-functions and weights that take the value $\psi = 0$ for all $r \geq r_0$. Examples are Hampel's two or three part redescending, Tukey's bisquare, Andrew's sine function e.t.c. (e.g. Hampel et al, 1986). The feature of redescending functions is that they apply a zero weight to all observations outside the influence region. This will, in general, work well if the initial guess is not extraordinarily distant from the expectation values. When however such an initial guess exists, the zero weight may tend to destroy desirable observations. Therefore, the most suitable ψ-functions are those that can extend their action to a few standard deviations of the (uncontaminated) error population. Although I have worked with, and I am satisfied that such ψ-functions, and their associated loss functions (e.g. Hampel's) work satisfactorily with most of the data I have studied, in the present I shall use, and advocate the use of the loss and influence function of Huber (1981), also discussed by Egbert and Booker (1986).

VI.1.1 Robust regression-M estimation

This loss function has the form

$$\rho(r) = \begin{cases} r^2/2, & |r| < r_0 \\ r_0 |r| - r^2/2, & |r| \geq r_0 \end{cases}$$

with $r_0 = \alpha\sigma$ (the standard deviation of the uncontaminated error distribution), and α is a real constant. This is a hybrid form that corresponds to L_2 minimization for the small residuals and L_1 minimization for the larger ones. The influence function is generically related to the loss function as

$$\psi(r) = \rho'(r) = \begin{cases} r , & |r| < r_0 \\ r_0 , & |r| \geq r_0 \end{cases}$$

with weights $w(r) = \psi(r)/|r|$, that modify the output observations according to the scheme (4.ii) above. The error scale $\sigma(k)$ is computed, for the n'th iteration as (Egbert and Booker, 1986)

$$\sigma(k) = \left[\frac{1}{\beta(2L - 4)} \sum_{n} r(n,k) \right]^{1/2}$$

i.e. is a scaled RMS error, with $\beta < 1$, which guarantees that the error scale is not underestimated during the successive iterations. β is a fuction of r_0. Egbert and Booker give a method to obtain the appropriate β for any choice or r_0, therefore I will not be concerned with this any more; more details can be found in their excellent paper. This ψ-function is not redescending. In fact, it is non-decreasing, which is a requirement for convergence to unique estimates, for the given $\rho(r)$. However, the weight function $w(\psi)$ is continuous, and can extend its influence over all r. Thus this ψ-function offers simplicity, and downweights outliers without break points other than r_0. This continuity is very attractive for most applications. General convergence is guaranteed, provided that (Huber, 1981)

$$\rho(0) = 0, \quad \rho'(r) > 0, \text{ and } 0 \leq \rho(r) \leq 1$$

More details can be found in the references cited.

The application of the robust iterative scheme described above, with $r_0 = 1.5$ and LS solutions to the successive approximations to $\hat{z}(k)$, converges, after several tens of iterations, to

$$z_{xx} = 22.01 \qquad z_{xy} = -97.19$$

$$z_{yy} = -67.29 \qquad z_{yx} = -7.73$$

which are not improvements that are worth the effort. The results shown are obtained by Golub's method, but the normal equations procedure

returns almost identical results. This failure does not discredit the robust estimation procedure. It rather shows that the wrong scheme has been used. One not quite often vocalized requirement for the application of such Robust LS (RLS) regression-M estimators is that the initial guesses should not be totally irrelevant to the expectation values of the parameters under estimation. If they are, or if a second, consistent noise population exists within the sample space, the influence function may destroy the adequate data populations as 'outliers'. Reasonable first estimates can be achieved if the non-Gaussian contaminants do not form consistent populations, as is the case with the Example Site 1. Also, the modification of the output vector **a** is not paired with equivalent modification of the input vector **b**. Although theoretically this makes no difference, in practice it can make the difference. The noisy waveforms in our case are not exclusive to the telluric fields; they exist in the magnetics as well, and their effects persist in the the formation of the covariance measure ($\mathbf{b}^T\mathbf{b}$), or the internal products in the Householder transformations. In such cases therefore, one may want to downweight the noise contributions in the input vector. This leads to variants of the regression-M estimators with implementation of WLS solutions.

If the error vector in (7) is assumed to contain uncorrelated entries, i.e. $r_i r_j = \delta_{ij}$, then the WLS solution can formally be defined so that $\mathrm{Var}\{a(n)\} = r^2(n)$, and $w(n) = 1/r^2(n)$, so that $\mathbf{W} = \mathrm{diag}\{w(n)\}$. This may, or may not, work. In this case it doesn't, for the same reasons that produce the failure of the LS solutions. If these are dominated by the consistent alien populations, then their residuals may actually be smaller than those of the adequate data. One quite smart choice, is to use the influence weights w(r) to weight the whole observational

equation, instead of modifying the output. The influence weight is clearly a function of location within the sample space. Thus, even if one extremely outlying observational equation is downweighted during the first pass, this will produce a cascading sequence of changes that may lead to improved estimation. This kind of weighting is actually equivalent to the standard regression-M estimator described above. In order to see how, consider that the minimum of the loss function $\rho(r)$ can be found by solving the system of equations

$$\sum_n \psi(r)\ \mathbf{b}(n) = \sum_n \rho'(r)\ \mathbf{b}(n) = 0$$

resulting from setting the derivative of the error variable (the loss function) equal to zero. This can be written as

$$\sum_n [\rho'(r)/r(n)]\ r(n)\ \mathbf{b}(n) = \sum_n w[r(n)]\ r(n)\ \mathbf{b}(n) = 0$$

which gives

$$\sum_n w[r(n)]\ [a(n) - \mathbf{b}(n)\ \mathbf{z}]\ \mathbf{b}(n) = 0,$$

i.e. is a form of WLS algorithm. Therefore the standard regression-M procedure and the iterative WLS scheme are equivalent, but their numerical efficiency can be very different. The variant of the robust WLS (RWLS) procedure I implement here, involves no modification of the output vector. Instead, at the begining of each iteration one sets up two vectors $\mathbf{a}' = \mathbf{a}$ and $\mathbf{b}' = \mathbf{b}$, and then, based on $\mathbf{a}$ and $\mathbf{b}$, proceeds to obtain the influence weights, as above. The weights operate on the primed vectors, from which one obtains the WLS solutions for $\hat{\mathbf{z}}(k)$. Thus, it can be seen that the procedure relies on the possibility of ever increasing residuals for the outlying observations, or equivalently, ever decreasing weight functions. The use of the auxiliary vectors $\mathbf{a}'$ and $\mathbf{b}'$ ensures that the weights do not affect the original observations, so that the outlying residuals can grow monotonically at each iteration. This is a rather crude form of the

RWLS scheme, but very effective. The procedure is guaranteed to converge under the conditions cited above. This will happen when the residuals stop growing significantly, i.e. when the optimal estimate <z> is reached.

The application of this RWLS scheme produces the following results:

GOLUB'S METHOD :	z_{xx} = 3.65	z_{xy} =-60.44	r_0 = 0.5
	z_{yx} =-31.36	z_{yy} = 4.32	
	z_{yx} =-41.79	z_{yy} = 5.27	r_0 = 1.0
	z_{xx} = 6.28	z_{xy} =-64.37	r_0 = 1.5
	z_{yx} =-68.83	z_{yy} = 7.94	
NORMAL EQUATIONS:	z_{xx} = 11.60	z_{xy} =-78.80	r_0 = 0.5
	z_{yx} =-61.43	z_{yy} = 7.20	
	z_{xx} = 15.64	z_{xy} =-87.21	r_0 = 1.0
	z_{yx} =-69.25	z_{yy} = 7.91	
	z_{xx} = 19.03	z_{xy} =-93.86	r_0 = 1.5

The study of these results reveals that spectacular improvements in the quality of estimation are achievable, but also reveals a number of other very interesting features. Firstly, note the impressive difference in the performance of the two numerical methods. The normal equation approach still suffers from the imprecisions of forming and inverting $\mathbf{b}^T\mathbf{b}$. The stability and versatility of the Householder transformation produces significantly more reliable results. Secondly, it is apparent that the quality of estimation essentially depends on the extent of the central and influence regions, i.e. the choice of r_0. For r_0 = 0.5 Golub's method produces almost optimal results. However, for $r_0 \geq 1$ it can clearly not cope. For Gaussian distributed errors, possibly contaminated by some outliers, a value of r_0 = 1.5 - 2.0 usually works very well. For very heavily contaminated (or multimodally distributed)

residuals, one has to extend the influence region, so as to ensure that outlying residuals will continue to grow monotonically. This is a somewhat distressing feature, because we will, in general, have no prior information about the distribution of the residuals and the optimal choice of r_0. If r_0 is fixed to small values (e.g. $r_0 = 1$), uncontaminated data will not be severly affected, and one can ensure reliable $\hat{z}$ estimates for most of the time. However, it is insidious to vary r_0 to very small values (< 1) as we do not want to heavily modify good data, and therefore it becomes insidious to base sensitive automatic selection schemes on optimistic expectations.

In conclusion, it appears that for all robustness introduced, the simple WA is still our best bet for an estimate of the expectation values of $\mathbf{z}$. A third (general) feature of the WLS solution is that it does not produce a measure of the z_{ij} population scatter, which is imperative for the proposed selection process. The error measure afforded by the LS algorithms is based on internal standard deviations, i.e. on the covariance measure $(\mathbf{b}^T\mathbf{b})$, which grossly underestimates the true dispersion of the solution vector element populations. Consider equation (7), left multiplied by $\mathbf{b}^T$, so that

$$\mathbf{b}^T\mathbf{a} = \mathbf{b}^T\mathbf{b}\ \mathbf{z} + \mathbf{b}^T\underline{\varepsilon}$$

This gives

$$(\mathbf{b}^T\mathbf{b})^{-1}\mathbf{b}^T\mathbf{a} = \mathbf{z} + (\mathbf{b}^T\mathbf{b})^{-1}(\mathbf{b}^T\underline{\varepsilon})$$

and can be interpreted to imply that

$$\hat{\mathbf{z}} = \mathbf{z} + (\mathbf{b}^T\mathbf{b})^{-1}(\mathbf{b}^T\underline{\varepsilon}) = \mathbf{z} + \mathbf{x}.$$

Therefore $\mathbf{x}$ is the measure of error in $\mathbf{z}$, whereupon the covariance of the solution vector $\hat{\mathbf{z}}$ is given by

$$\mathrm{cov}\{\hat{\mathbf{z}}\} = \mathbf{x}\ \mathbf{x}^T .$$

By considering the magnitude of the quantities involved in this

expression, it can easily be seen that this will underestimate the actual adequate data scatter. I think I can provide a solution to both the problems of expectation value and scatter estimation, by using one robust alternative to the LS iterative schemes. This is yet another variant of the regression-M estimator, which involves computation of the elements z_{ij} of the solution vector z with W-estimators.

VI.1.2 Robust W-estimates of location

The W-estimators are defined as a weighted average <z> of the observations

$$z(L)\{z_1,z_2,\ldots,z_L\} = \sum_1 w(l)\, z(l) \left[\sum_1 w(l) \right]^{-1}$$

where the weights depend on the observations through

$$w(l) = w[z(l) - \langle z \rangle]$$

i.e. they are a function of the location of z(l) within the sample space. The W-estimator therefore, satisfies the celebrated WA equation

$$\langle z \rangle = \sum_1 w[z(l) - \langle z \rangle]\, z(l) \left[\sum_1 w[z(l) - \langle z \rangle] \right]^{-1}$$

This can be modified to yield

$$0 = \sum_1 w[z(l) - \langle z \rangle]\, [z(l) - \langle z \rangle] \left[\sum_1 w[z(l) - \langle z \rangle] \right]^{-1}$$

which implies that

$$0 = \sum_1 \rho[z(l) - \langle z \rangle]$$

with

$$\rho'[z(l) - \langle z \rangle] = [z(l) - \langle z \rangle]\, w[z(l) - \langle z \rangle].$$

Note that this is just another expression for the general form of a loss function (the error scale is ommited), since <z> = E{z(l)}. Therefore, the (iterated) W-estimator is actually a variant of the the M- type estimators of location, and as such, they posses the same influence function and asymptotic variance (Hampel et al, 1986). <z> is usually

determined iteratively, starting from an initial guess $\langle z_0 \rangle$ (e.g. the mean, or the median), and computing

$$\langle z(k+1) \rangle = \sum_{l} w[z(l) - \langle z(k) \rangle] \, z(l) \left[\sum_{l} w[z(l) - \langle z(k) \rangle] \right]^{-1}$$

(iteratively reweighted least squares) until convergence.

Now, consider that in cases where the number of equations (constraints) N >> M, the number of unknowns, as is usually the case with AMT data analysis, one can always solve N/M systems of order MxM, to yield an ensemble of N/M estimates of the solution vector, each one of them exactly satisfying one of the MxM systems. In our case, this corresponds to N=2L equations in M=2 unknowns, and therefore the L resulting estimates z(l) correspond to each data window, solved by using (4). Instead of implementing the usual LS algorithms, we can use the ensemble of z(l) to compute the successive approximations to the expectation value $\langle z \rangle$. The estimation of $\langle \hat{\mathbf{z}}(k) \rangle$ can be performed with the *one-step* W-estimator. In this case, the iterative scheme will operate as follows:

(1) Obtain an initial estimate $\hat{\mathbf{z}}(0)$ for the solution vector **z** and the error scale $\hat{\sigma}(0)$. This can be any reasonable estimator, i.e. the median, weighted median, RWLS solutions, or more likely the mean or WA first approximations. Then, for k=0,....

(2) Compute the predicted output observation $\hat{a}(n,k) = \mathbf{b}(n)\, \mathbf{z}(k)$ and the residual $r(n,k) = a(n,k) - \hat{a}(n,k)$,

(3) Modify the output observation as $a(n,k+1) = \hat{a}(n,k) + w(r)\, r(n,k)$

(4) Use the modified observations a(n,k+1) and a(n+1,k+1) to compute L estimates $\hat{\mathbf{z}}(l,k+1)$ of the solution vector by using (4).

(5) Form the W-estimators

$$\langle z_{ij}(k+1) \rangle = \sum_{l} z_{ij}(l,k+1)\, w(l,k+1) \left[\sum_{l} w(l,k+1) \right]^{-1}$$

Also, compute the standard deviation s(k+1) of the $\hat{z}(1,k+1)$ populations and the new rms error scale $\hat{\sigma}(k+1)$.

Then iterate from (2).

The weight factors for the W-estimator require some consideration. We desire that our ideal weight disappears when it is not needed any more, but we also require that it does *not*, when $\hat{z}(1,k)$ is not fully recovered. Truly, there may exist noise waveforms producing small residuals, that are not affected by the process of correction. Alternatively, the process of correction ceases when one outlying residual drops within the central region, but this doesn't mean that this estimate is fully recovered. Moreover, we want our weight to retain the power of eliminating bad data points, other than outliers. Therefore, the optimal weight will behave in a manner that

(i) retains its power for data other than outliers,

(ii) vanishes for well recovered data, and

(iii) can perceive and downweight noise that evades the correcting action of the robust processor.

The ideal weight is very difficult, if not impossible to obtain. One obvious choice is to use the influence weight w(r), which has the functional form

$$w\{r(n,k)\} = f\{a(n,k) - \mathbf{b}(n)\hat{\mathbf{z}}(k)\} = f\{\mathbf{b}(n)[\hat{\mathbf{z}}(1,k) - \hat{\mathbf{z}}(k)]\}$$

and therefore is a genuine function of the location of $\hat{\mathbf{z}}(k)$. Since the influence weight depends on the vector $\hat{\mathbf{z}}(k)$, it can be used on both of its elements. By choosing

$$w(1,k+1) = \min\{w[r(n,k)/\ \hat{\sigma}(k)],\ w[r(n+1,k)/\ \hat{\sigma}(k)]\}$$

we ensure maximum weighting against the outlier producers. This choice of weighting has its merits, but also disappears as soon as an estimate

$\hat{\mathbf{z}}(k)$ produces a residual r(l,k) within the central region. This will not always be desirable. Thus, w(l,k) cannot attain properties (i) and (iii). On the other hand, this type of weight function is very stable and reliable, and will always ensure that the procedure converges satisfactorily; of the three types of weight functions I have investigated, this the only type that can be used as a standard, for all data sets.

A second candidate approximation to the ideal weight is the use of the absolute deviation $\Delta z = |\hat{\mathbf{z}}(l,k) - \hat{\mathbf{z}}(k)|$, so that $w(l,k) = 1/\Delta z$. This weight can be effective in the initial stages of the iterative procedure, but it becomes unstable as the process converges, or when the deviation Δz happens to be very small. If it is implemented, it is desirable that it be normalized to unity. Also, care must be taken so that it ceases to operate when Δz becomes too small. A third approximation to w(l,k), can be afforded if we reconsider the properties of the random error, and specifically its dependence on the auto-covariance (auto-power) of the output electric field. Consider equation (3), which is derived for a single data window. In order to comply with the definitions of the present section, this can be rewriten as

$$\mathrm{cov}\{e_i(l)e_i(l)\} = \sum_{t=0}^{N} \{e_i(l,t) - \mathbf{h}(l,t)\ \mathbf{z}(l)\}\ ,$$

where $\mathbf{h}(l,t) = [h_x(l,t)\ h_y(l,t)]$. We have seen that this expression leads to equation (5), which predicts the auto-covariance (auto-power) of the electric field, based on $\mathbf{z}(l)$. If we substitute $\hat{\mathbf{z}}(l,k)$ for $\mathbf{z}(l)$, we can immediately afford a measure of discrepancy between the observed auto-covariance, and that predicted by our current (k'th) estimate for the l'th data window. Moreover, if we substitute $\hat{\mathbf{z}}(k)$ for $\mathbf{z}(l)$, we can afford a measure of the deviation of the observed

auto-covariance, and that predicted by our current (k'th) best estimate for the expectation values $\langle z \rangle$. By using this simple trick, we can predict, modify, and correct for the erring E-field auto-covariances. The correction can simply assume the form of the equivalent corrections for the **a** vector. Thus, the robust W-estimator scheme can be augmented, for the k'th iteration, as follows:

(1) (i) Predict and modify the **a** vector in the prescribed manner.

(ii) Simultaneously, based on $\langle \hat{z}(k) \rangle$, predict the E-field auto-covariance $\hat{p}(1,k)$ by using (5), and the residual $x(1,k) = p(1,k) - \hat{p}(1,k)$; modify p in the prescribed manner, i.e.

$$p(1,k+1) = \hat{p}(1,k) + x(1,k)\, w\{x(1,k)/\hat{\sigma}_p(k)\}$$

(2) Compute new estimates $\hat{z}(1,k+1)$, based on the modified observations $a(n,k+1)$ and $a(n+1,k+1)$; simultaneously, use $p(1,k+1)$ to obtain the (modified) random errors $\hat{\varepsilon}(1,k+1)$.

(3) Compute the W-estimators for $\hat{z}(k+1)$, using $\hat{\varepsilon}(1,k+1)$ weighting. Compute the standard deviation of $\hat{z}(1,k+1)$, and new error scales for both $a(1,k+1)$ and $p(1,k+1)$. Then resume from (1).

The modified E-field auto-covariances are be used to obtain new random errors, that correspond to the new estimates $\hat{z}(1,k+1)$. In this way, the error is also transformed into a function of the location of $z(1,k)$ with respect to $\langle \hat{z}(k) \rangle$. The standard expression for the random error, is a function of $p(1,k)$, the output E-field auto-covariance. This in turn, is a function of z. Thus one can write

$$\varepsilon_r = f\{\mathrm{cov}\{e_i e_i\}\} = f\{z\}$$

All the other parameters, are observables, or depend on $p(1,k)$ (e.g. R_i^2). If we replace the observed auto-covariance with the modified auto-covariance, we have

$$\hat{\varepsilon}(1,k) = f[\hat{z}(1,k)],$$

and the change in the error between the k'th and k+1'th iteration is a function of the change in $\hat{\mathbf{z}}(1,k)$. We can write

$$\Delta\varepsilon(1) = \hat{\varepsilon}(1,k+1) - \hat{\varepsilon}(1,k) = f\{\hat{\mathbf{z}}(1,k+1) - \hat{\mathbf{z}}(1,k)\} = f\{\Delta\mathbf{z}(1)\}$$

However, because we have

$$\Delta\mathbf{z}(1) = f\{\hat{\mathbf{a}}(n,k) - \mathbf{b}(n)\,\hat{\mathbf{z}}(k)\} = f\{\mathbf{b}(n)[\hat{\mathbf{z}}(1,k) - \hat{\mathbf{z}}(k)]\}$$

we finally conclude that

$$\Delta\varepsilon(1) = f\{\mathbf{b}(n)[\hat{\mathbf{z}}(1,k) - \hat{\mathbf{z}}(k)]\},$$

i.e., the degree of change in the error is ultimately associated with the degree of change in the estimation of $\mathbf{z}$. The iterative procedure can be thought of, as minimizing the loss function, conditionally on minimizing the output auto-covariance (or the expression for the random error). In conclusion, the modified random error is a legitimate weight factor. However, because it is subject to the modification of the output signal auto-covariance, it can become quite unstable, and a note of serious caution applies with respect to its usage.

The application of the procedure gives, with $r_0 = 1.5$,

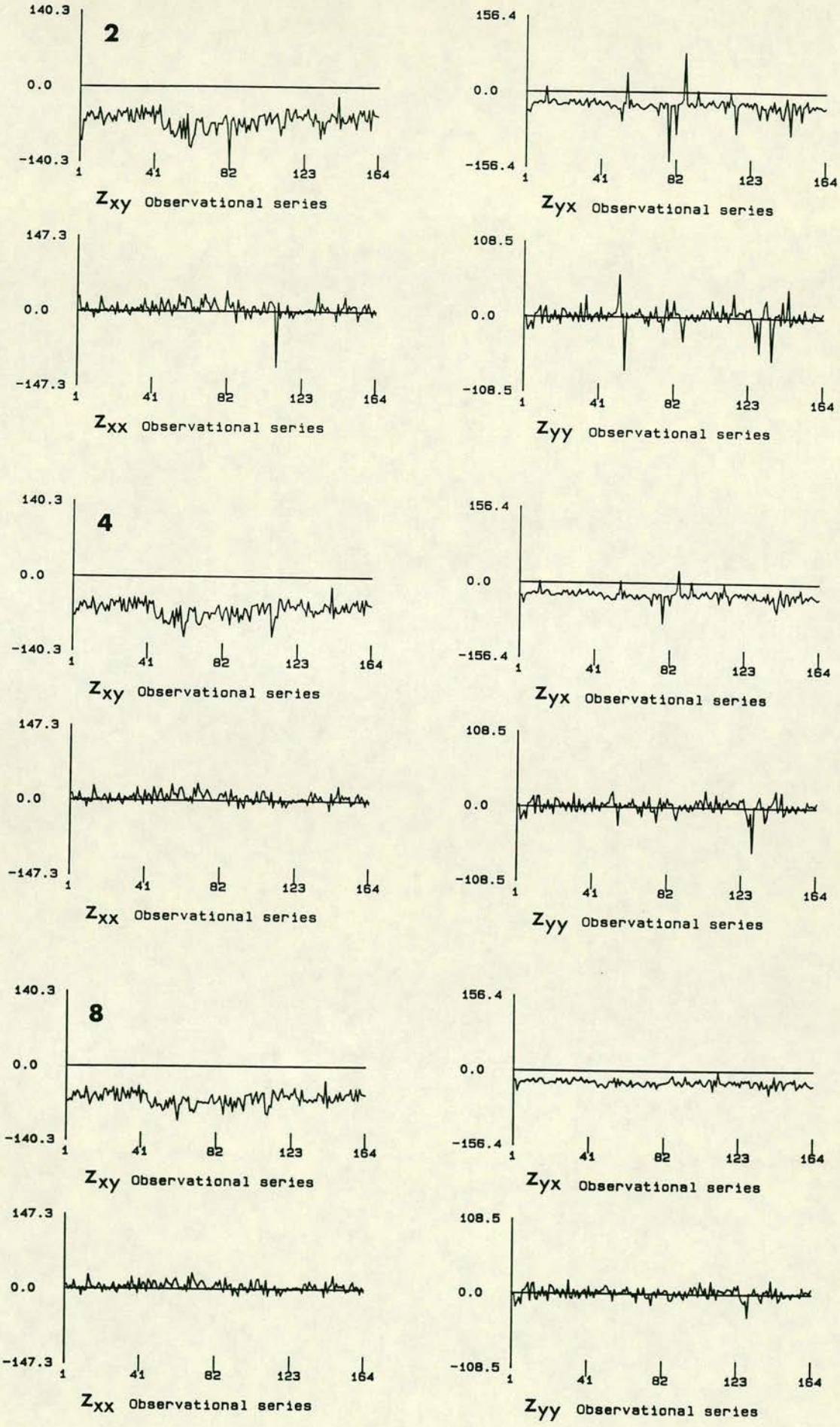
$$z_{xx} = 2.29 \pm 8.98 \qquad z_{xy} = -58.84 \pm 12.48$$

$$z_{yx} = -24.25 \pm 5.99 \qquad z_{yy} = 0.96 \pm 8.27$$

i.e. improves on the WA initial guess, and spectacularly reduces the population scatter. The operation of procedure can be seen in Figure 9a, where the results of the 2'nd, 4'th and 8'th iteration are shown. The modification of the output observations causes the outliers to be pulled towards the expectation values. Eventually, most of them will vanish, so that the resulting populations can yield an optimal (B-robust) mean and standard deviation. The results presented above have been obtained with modified random error weighting, which, for the data in question, was found to be a stable and well behaved function. The modified random errors for the 2'nd, 4'th and 8'th iterations, can

FIGURE 9 Robust cleaning of the contaminated z(0) factors of Example Site 1. **(a)** are the recovered (cleaned) z_{ij} observational series after 2,4 and 8 iterations; **(b)** are the corresponding modified random errors ε_{ij}, used as weights.

(a)


2
140.3
0.0
-140.3
1
41
82
123
164
Zxy Observational series
156.4
0.0
-156.4
Zyx Observational series
147.3
-147.3
Zxx Observational series
108.5
-108.5
Zyy Observational series
4
8

(b)

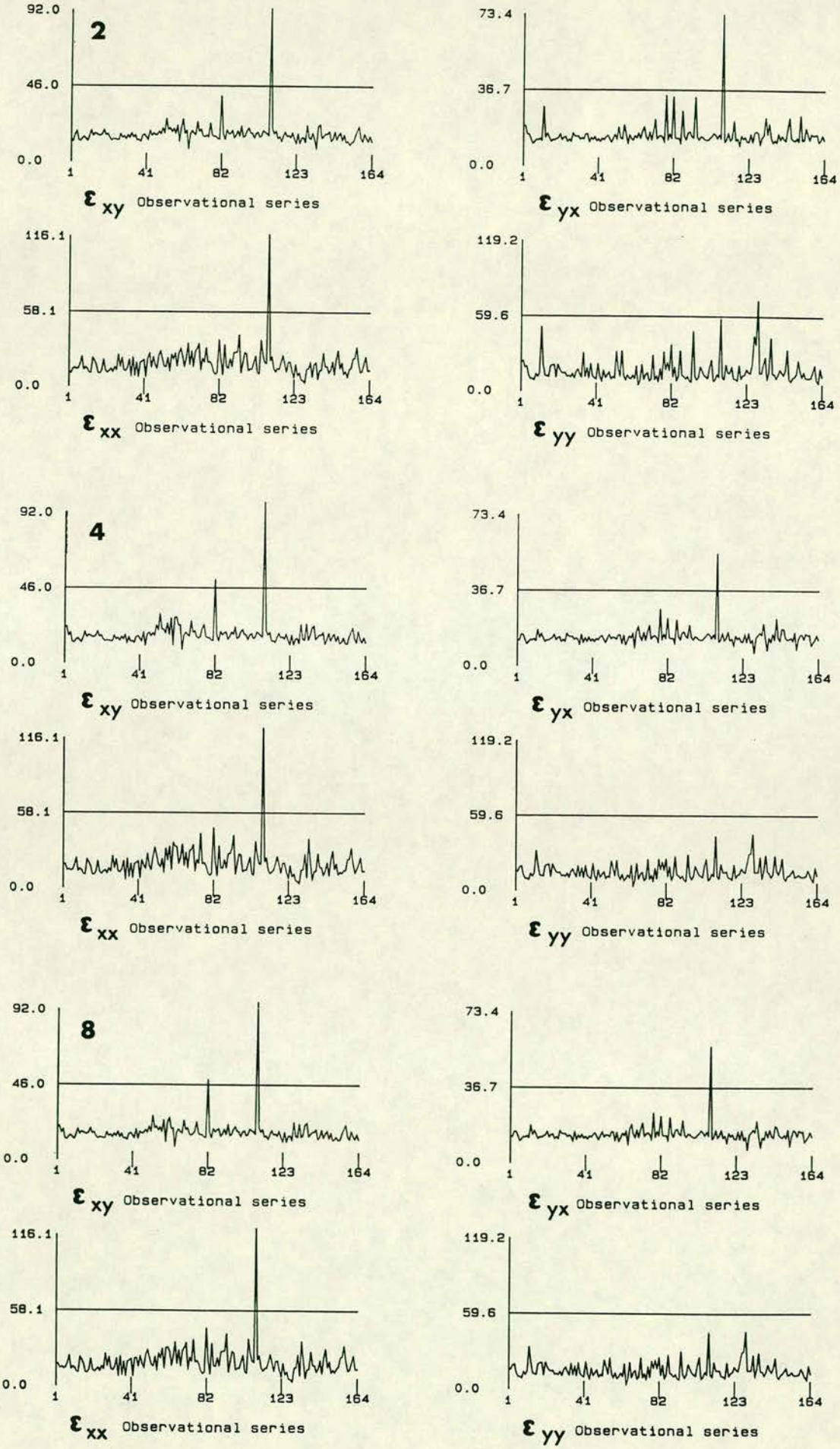


FIGURE 9 ..continued

be seen in Figure 9b.

The procedure, being equivalent to the standard regression-M estimation, is covered by similar theory, and will always converge under the conditions given above. It actually locates the optimal $\langle\hat{z}\rangle$ quite rapidly, and then takes its time to converge to the n'th decimal point accuracy. If such an accuracy is not desired, (in this application it certainly is useless), it is usually safe to truncate the process after 10-15 iterations. This scheme has the obvious advantage of avoiding all the numerical complications encountered in the usual LS procedures. Thus, powerful events remain isolated outliers and do not disproportionately affect estimation. Moreover, it is quite simple to program and operate, even in small computer systems.

There exists an important point that requires some special consideration. The standard deviation of the corrected z_{ij} populations, as computed with the robust W-estimator scheme, is a function of the choice of r_0. If the central region shrinks too much, the corrective action will extend its influence over the majority of the observational equations, and 'kill' everything within its reach, even just fractions of the uncontaminated data one s.d. away. Thus, for $r_0 < 1$, the resulting s.d. is usually underestimated and unrealistic. Likewise, for large r_0 (e.g. > 3) the rigour of the correction is relaxed, and the s.d. may be overestimated. As mentioned previously, for Gaussian residuals, contaminated with a few extreme outliers, a value of $r_0 = 1.5\text{-}2$ works well. The value I have used is $r_0 = 1.5$, and is a result of long standing wisdom of other investigators; it appears to work very well with this data set too. Note that the extent of the central region is related to the s.d. of the true (uncontaminated by non-Gaussian

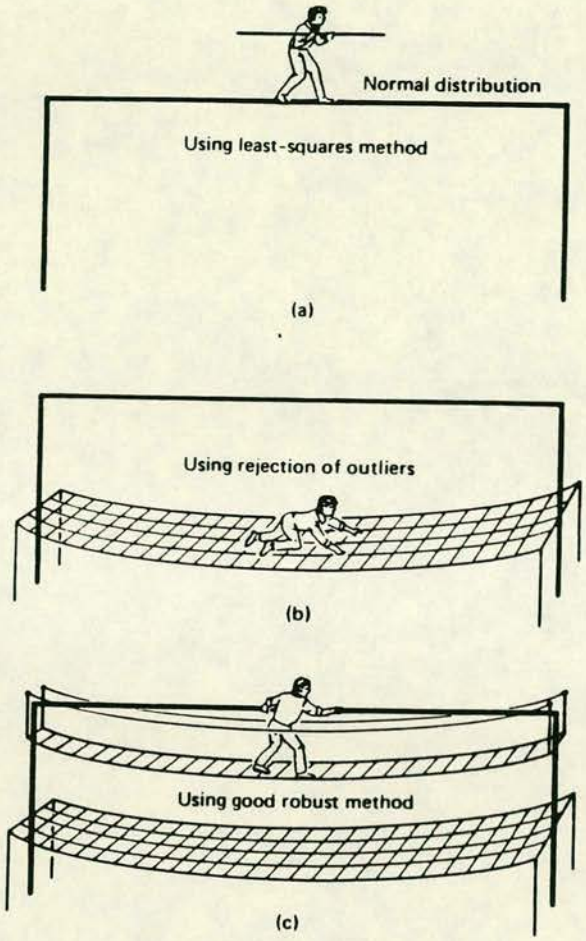


FIGURE 10 Various ways of analyzing data. Reprinted from Hampel et al (1986).

noise) data population. By keeping r_0 in the order of 1.5-2 s.d. of the Gaussian residuals, one translates similar constraints to the resulting ensemble of the solutions $\hat{z}(l,k)$. Thus, the computed s.d. is neither grossly over- or under- estimated with respect to the expectation. It is still not true, but it is realistic and practical, at the level of the robustness attainable by the procedure. The procedure can also be extended to *Frequency Domain* estimation, with apparent advantages in terms of speed and efficiency. Details will be discussed in the next section.

As a conclusion, of this brief study on estimation procedures, I would like to point out that there exists a scheme that is both robust and versatile for a given application. For the problem of the z(0) factors (or indeed $Z(\omega)$ in the frequency domain), I believe that the robust W-estimator accomplishes both tasks, even in the presence of very heavy noise contamination. In order to avoid any more lengthy discussions, I prefer to refer the reader to Figure 10, reprinted from Hampel et al (1986), which clearly explains everything.

VII. Standard Frequency Domain estimation techniques.

The frequency domain LS solutions of Sims et al (1971) are derived under the assumption that noise terms reside in the output channels only. This -artificial- condition introduces bias in the estimated impedances due to noise in the input channels. Bias is upward or downward, depending on the solution. Define the admittance tensor $\mathbf{Y}(\omega)$ so that

$$\mathbf{H}(\omega) = \mathbf{Y}(\omega)\mathbf{E}(\omega), \qquad \mathbf{Z}(\omega) = \mathbf{Y}^{-1}(\omega)$$

with $\mathbf{H}=[H_x\ H_y]$, $\mathbf{E}=[E_x\ E_y]$, and rewrite the 2-input 1-output relationship

as

$$\langle 1 \rangle = A\langle 2 \rangle + B\langle 3 \rangle \qquad (9)$$

The ω dependence will be henceforth dropped for simplicity. The LS solutions of Sims et al (1971) for B will be of the type

$$B = \frac{S_{22}S_{31} - S_{32}S_{21}}{S_{22}S_{33} - S_{23}S_{32}}$$

where for example S_{23} is the observed cross-spectrum between the input channels. If $\langle 1 \rangle = E_x$, $\langle 2 \rangle = H_x$ and $\langle 3 \rangle = H_y$, (9) will return Z_{xy} biased downwards by noise in **H**. If $\langle 1 \rangle = H_x$, $\langle 2 \rangle = E_x$ and $\langle 3 \rangle = E_y$, (9) will yield Y_{xy} biased downwards by noise in **E**; likewise, all the elements in **Y** will be biased downwards, so that $\mathbf{Y}^{-1}$ will be biased *upwards*. With no a priori information about the distribution of noise terms across the data channels, a common practice is to utilise (9) to provide limiting bias cases. 'Unbiased' estimates of the impedance tensor can then be computed from the upper and lower bounds set by **Z** and $\mathbf{Y}^{-1}$. The actual estimates of these matrices will take the form of averages across the sample space provided by the recorded data realisations (windows). One very popular method to obtain those averages is to define as sample space the spectra of all the data realisations, and involves *spectral stacking* in order to derive the required response function estimates from the average auto- and cross-spectral matrix. This approach however is liable to the possibility that the spectral matrix be dominated by a few very powerful data realisations, with unpleasant results. An alternative approach is to define as sample space the ensemble of response functions associated with each data realisation, and, use them to derive a WA of the response functions by expressions of the type given by (8). The use of partial coherences as weights has been reported by Jones et al (1983). This may be disastrous in cases of coherent noise. The author has found that the random errors of Pedersen

(1982) work very well. The isomorphism between the first order time domain and frequency domain processes suffices to explain why; the reasons are identical to those described above. However, because of the apparent importance of the quantity as a diagnostic aid and weight factor, a more detailed study is provided in Appendix 4.B. Note that for a proper implementation of the WA procedure the real and imaginary parts of $\mathbf{Z}(\omega)$ must be treated *independently*, because they constitute statistically independent, Wishart distributed quantities. Application of the WA procedure to quantities such as apparent resistivity and phase is not warranted, because they constitute non-linear transforms of the complex impedance and, therefore, their statistical distribution is *unknown*.

Both approaches assume normal (Gaussian) properties of the sample space in the frequency domain. Note however, that although derived from the same auto- and cross-spectra, $\mathbf{Z}$ and $\mathbf{Y}^{-1}$ should be treated independently since they represent biased estimates of the true expectation values. Bias reduction across the sample space takes place by winnowing. In its simplest (and most popular) form this constitutes rejection of those realisations, for which the multiple (predicted) or partial coherences fall bellow a preset threshold. In theory, if the data set provides an adequate distribution of high multiple coherence functions (e.g.>0.80) bias errors are reduced to the level of random errors. The 'true' expectation value of the impedance tensor can be estimated with a number of different approaches (e.g. Adam and Vero 1976, Jones et al 1983, Cox et al 1980). In this study the weighted average

$$\langle T_{ij} \rangle = \frac{Z_{ij}(\Delta Z_{ij})^{-2} + Y_{ij}^{-1}(\Delta Y_{ij}^{-1})^{-2}}{(\Delta Z_{ij})^{-2} + (\Delta Y_{ij}^{-1})^{-2}}$$

is adopted.

VII.1 Robust Frequency Domain estimation.

The concepts of the time domain robust estimation of the z(0) factor populations, can readily be extended to frequency domain processing; here there exists a choice of systems to solve. We can set up the frequency dependent system of observational equations

$$\mathbf{A}(\omega) = \mathbf{B}(\omega)\ \mathbf{Z}(\omega) \tag{10}$$

or explicitly,

$$\begin{bmatrix} \mathrm{ReE}_{i_1} \\ \mathrm{ImE}_{i_1} \\ \mathrm{ReE}_{i_2} \\ \mathrm{ImE}_{i_2} \\ \cdots \\ \mathrm{ReE}_{i_L} \\ \mathrm{ImE}_{i_L} \end{bmatrix} = \begin{bmatrix} \mathrm{ReH}_{x_1} & -\mathrm{ImH}_{x_1} & \mathrm{ReH}_{y_1} & -\mathrm{ImH}_{y_1} \\ \mathrm{ImH}_{x_1} & \mathrm{ReH}_{x_1} & \mathrm{ImH}_{y_1} & \mathrm{ReH}_{y_1} \\ \mathrm{ReH}_{x_2} & -\mathrm{ImH}_{x_2} & \mathrm{ReH}_{y_2} & -\mathrm{ImH}_{y_2} \\ \mathrm{ImH}_{x_2} & \mathrm{ReH}_{x_2} & \mathrm{ImH}_{y_2} & \mathrm{ReH}_{y_2} \\ & \cdots\cdots\cdots\cdots\cdots\cdots & & \\ \mathrm{ReH}_{x_L} & -\mathrm{ImH}_{x_L} & \mathrm{ReH}_{y_L} & -\mathrm{ImH}_{y_L} \\ \mathrm{ImH}_{x_L} & \mathrm{ReH}_{x_L} & \mathrm{ImH}_{y_L} & \mathrm{ReH}_{y_L} \end{bmatrix} \begin{bmatrix} \mathrm{Re}(Z_{ix}) \\ \mathrm{Im}(Z_{ix}) \\ \mathrm{Re}(Z_{iy}) \\ \mathrm{Im}(Z_{iy}) \end{bmatrix} \tag{10a}$$

with i=x,y, which has the same form with that considered by Egbert and Booker (1986). Alternatively, we can generate the direct extension of (7) in the frequency domain, with the vectors **A**, and **B** now containing the auto and cross-spectra of the electric and magnetic fields, i.e.

$$\begin{bmatrix} \mathrm{Re}(E_i H_x^*)_1 \\ \mathrm{Im}(E_i H_x^*)_1 \\ \mathrm{Re}(E_i H_y^*)_1 \\ \mathrm{Im}(E_i H_y^*)_1 \\ \cdots \\ \mathrm{Re}(E_i H_x^*)_L \\ \mathrm{Im}(E_i H_x^*)_L \\ \mathrm{Re}(E_i H_y^*)_L \\ \mathrm{Im}(E_i H_y^*)_L \end{bmatrix} = \begin{bmatrix} \mathrm{Re}(H_x H_x^*)_1 & 0 & \mathrm{Re}(H_y H_x^*)_1 & -\mathrm{Im}(H_y H_x^*)_1 \\ 0 & \mathrm{Re}(H_x H_x^*)_1 & \mathrm{Im}(H_y H_x^*)_1 & \mathrm{Re}(H_y H_x^*)_1 \\ \mathrm{Re}(H_x H_y^*)_1 & -\mathrm{Im}(H_x H_y^*)_1 & \mathrm{Re}(H_y H_y^*)_1 & 0 \\ \mathrm{Im}(H_x H_y^*)_1 & \mathrm{Re}(H_x H_y^*)_1 & 0 & \mathrm{Re}(H_y H_y^*)_1 \\ & \cdots\cdots\cdots\cdots\cdots\cdots & & \\ \mathrm{Re}(H_x H_x^*)_L & 0 & \mathrm{Re}(H_y H_x^*)_L & -\mathrm{Im}(H_y H_x^*)_L \\ 0 & \mathrm{Re}(H_x H_x^*)_L & \mathrm{Im}(H_y H_x^*)_L & \mathrm{Re}(H_y H_x^*)_L \\ \mathrm{Re}(H_x H_y^*)_L & -\mathrm{Im}(H_x H_y^*)_L & \mathrm{Re}(H_y H_y^*)_L & 0 \\ \mathrm{Im}(H_x H_y^*)_L & \mathrm{Re}(H_x H_y^*)_L & 0 & \mathrm{Re}(H_y H_y^*)_L \end{bmatrix} \begin{bmatrix} \mathrm{Re}(Z_{ix}) \\ \mathrm{Im}(Z_{ix}) \\ \mathrm{Re}(Z_{iy}) \\ \mathrm{Im}(Z_{iy}) \end{bmatrix} \tag{10b}$$

with i=x,y. The form (10a) is clearly neater, and more manageable, and

therefore more suitable for LS and RLS algorithms. Note however that it produces a twofold increase in the size of the input matrices, with all the important repercussions this has, in the degree of computational accuracy needed to solve the system. An additional problem that may be encountered in frequency domain robust estimation of AMT data, is that the auto and cross spectra of the EM field components comprise very small numbers, since there exists a very small amount of power in the fields. This may cause significant problems with numerical accuracy; double precision arithmetic will not necessarilly bail us out, and, in addition, the matrix $\mathbf{B}(\omega)$ will be unstable and prone to becoming singular. Such problems can be avoided by either rescaling the problem, or, considering the second form (10b), which is more suitable for the robust W-estimation procedure. It is, also, the only form applicable when reference channels exist, and one wants to extend the robust procedure to remote reference applications. In the second case we do not actually need to set up the system vectors, since the W-estimator can be made complex, provided that we use real weights. In addition, all the other operations required by the estimation procedure (i.e. predictions) can assume the form of straight complex multiplications. Therefore, the iterative solution of the system is not as difficult as it first appears to be. It is important to note that in this case we have 4L equations in 4 unknowns, because the real and imaginary parts of any complex quantity are typically taken to be statistically independent and uncorrelated. Thus, in theory at least, the output vector modifications need to be carried out on the real and imaginary parts separately, with independent weights. Thus, the only difference between the time domain and frequency domain W-estimators is the conversion of

the necessary multiplications to complex, and the solution for the Z(l,k) impedance determination (data window), with the classical LS method of Sims et al (1971).

The following details may be significant, and vital when applying the robust procedure to AMT data. As stated above, the spectral matrices comprise very small numbers. The AMT estimation problem is ill-posed, i.e. small changes in the observed field values may lead to appreciable changes in the estimated response functions. Therefore, when we impose modifications, we may be asking for trouble if they are not done carefully! This problem will, in general, be less serious when using LS algorithms, but it will be accentuated in the case of the W-estimator approach, because the iterative procedure requires the calculation of response function from single data windows. Thus, differential modification of the real and imaginary parts of the output cross-spectra may lead to *temporarily* bizarre (unstable) values of the $\hat{Z}(l,k)$ estimate; the situation will worsen with increasing noise contamination. I have found two ways to overcome this problem. One way is to define a narrow central region (i.e. $r_0 = 1$) and let the process converge (it always will), after a *large* number of iterations. Alternatively, one may decide to bent the rule of statistical independence of the real and imaginary parts, and apply the *same* weight to modify all entries $Re(E_i H_x^*)$, $Im(E_i H_x^*)$, $Re(E_i H_y^*)$ and $Im(E_i H_y^*)$ of the output vector. The second procedure converges considerably faster, but the results of both remedies are very comparable.

Robust Frequency Domain estimation can be very useful in cases of data contaminated as shown, but it is a very slow and tedious process; it

must be repeated over all data windows, and over all spectral estimates. The time domain selection procedure was designed so that, for AMT data, this complication can be avoided. In my experience, after careful time domain data selection with the procedures described above, it generally becomes unnecessary. Then, the frequency domain WA is a most successful estimator of the expectation values, and the data scatter is trully realistic (see examples below). Nonetheless, because I cannot say that I have exhaustively investigated all kinds of data, I will not commit a statement. As with all statistical procedures, there are no guarantees, only good chances. The scientist's problem therefore reduces into trying to make his chances very good, and this task is possible with the time domain selection procedure. Robust Frequency Domain estimation is most certainly useful for long period (MT) data, since the time domain data selection method is not always applicable for long period processes (see discussion below). However, because I limit my present investigations to AMT data, I shall not proceed to considerations of MT data robustification.

The time and frequency domain robust procedures, perform almost identically, and, nothing is to be gained with detailed numerical examples and comparisons. The examples given for the z(0) factor estimation can be used as a guide for the frequency domain estimation, under similar noise environments.

VIII. Application of the Time domain data selection procedure to Frequency Domain data processing.

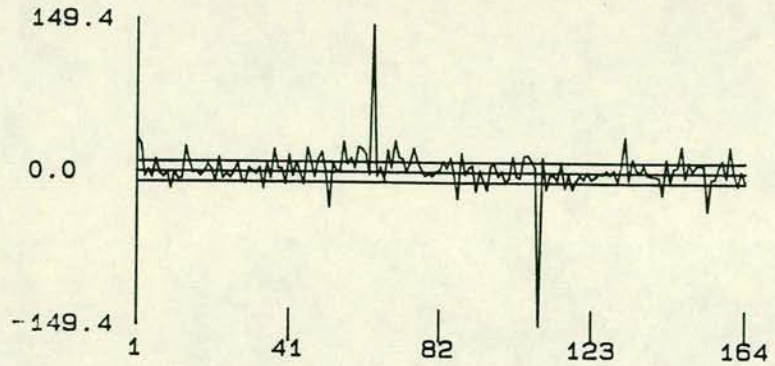
VIII.1 Example Site 1

The observational series, as well as the histograms for the z(0) factors for this site have been thoroughly examined above. Now, consider Figure 12a, which shows the 'unbiased' ($\langle T_{ij} \rangle$) estimate from the same sample space, computed with the two frequency domain methods discussed above, with no time domain selection. The MAXENT spectral analysis method has been used. The thick line depicts the WA result, while the thin line corresponds to the average spectral matrix, resulting from spectral stacking. Errors are not shown. Spectral stacking involved eight independent stacks, one for each of the elements of the Z and Y^{-1} matrices, subject to predicted (>0.80) and partial (>0.60) coherence thresholds. These are very strict criteria and would theoretically minimize bias by allowing only the best S/N data windows through. Identical coherence tests were required for the weighted averages as well. It is clear that the 'ensemble averages' are inconsistent between the two methods, especially for the Z_{yx} estimates. In this case we know why. The transient stray currents are much more numerous and powerful in the E-W direction (e.g. Figure 1) and their coherent contributions (multiple coherent noise of Kröger et al) tend to dominate the spectral accumulations. The resulting 'stacked' impedance is therefore shifted upwards. The overall shape of the function is not greatly affected; the noise realisations that passed the coherence tests were not sufficiently numerous or large, so as to inflict severe damage of this kind, i.e. frequency dependent distortion. Moreover, the WA function is considerably smoother than the 'stacked' one. Figure 12b shows the WA function with its associated standard deviation. Although here we have a smoother and more reliable expectation value, the scatter associated with inconsistent noise sources is still there, accentuated by the cooperation of the inadequate ones. In such a high resolution

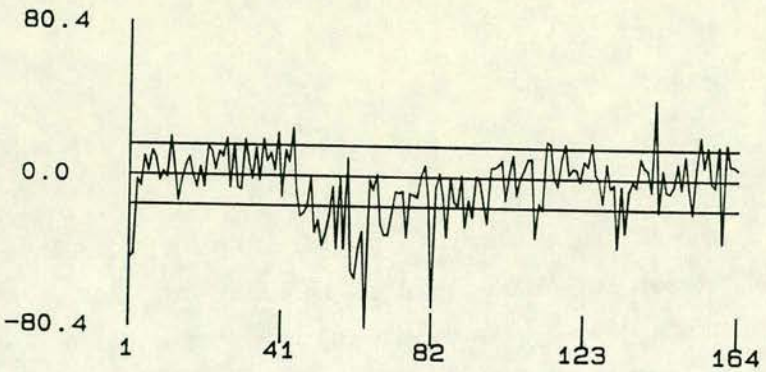
function the variation of the error shows exactly where the most serious noise spectral lines are located (note that the 50Hz noise harmonics have been removed).

The application of the proposed data selection technique can follow either scheme, automatic or manual. In this case study the available information is concise and clear; it allows an equally efficient manual selection of the permissible bounds. This procedure also enables the selection of data populations with specific characteristics for further study. Note however that for selection scheme (B), a *finer* class interval than the one presented for demonstration is recommended. Experimentation with either scheme will lead to almost identical results. The automatic selection procedure, employing Robust W-estimators and no R^2 thresholds, allowed only 69 data windows to pass the constraints. The action of the selection scheme is depicted in Figure 11. The observational series shown here correspond to the differences $\langle z \rangle - z(1)$, where $\langle z \rangle$ refers to the robust W-estimate of the expectation values. The bounded region is three (robust) standard deviations wide. With acceptance based on any combination of z_{ij}, only the data windows with z(0) factor solutions inside the bounds would satisfy the criteria. The frequency domain WA results are shown in Figure 12c, produced as per Figure 12a. No comparison between the 'stacked' and 'averaged' estimates is given because they become *absolutely consistent*. The reduction of scatter of the WA is dramatic. More important however is the fact that no local variation exists in the error; it now becomes a smooth function of frequency. This indicates that the data realisations picked up by the selection technique form consistent, high S/N ratio populations. Finally, in Figure 13, I

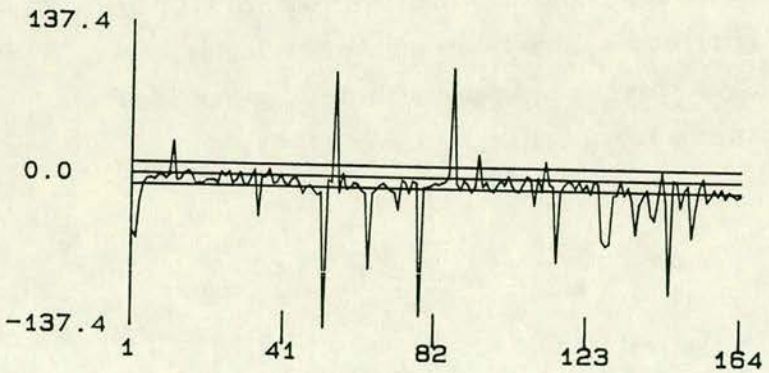
FIGURE 11 A visualization of the operation of the data selection procedure, for the data windows (z(0) factors) of Example Site 1. The robustly computed location parameter $\langle z \rangle$ has been removed. The width of the bounded region is three (robust) standard deviations. Only data windows with z(0) factors within the bounded regions are acceptable.



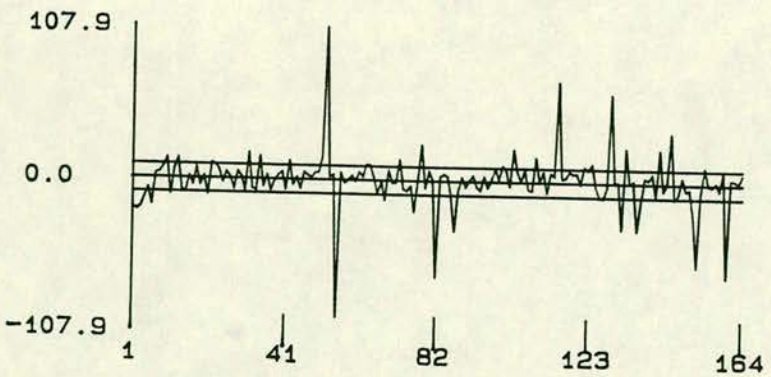
Z_{xx} Observational series



Z_{xy} Observational series

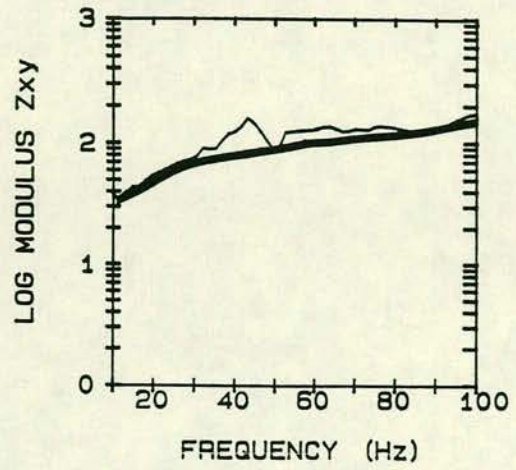


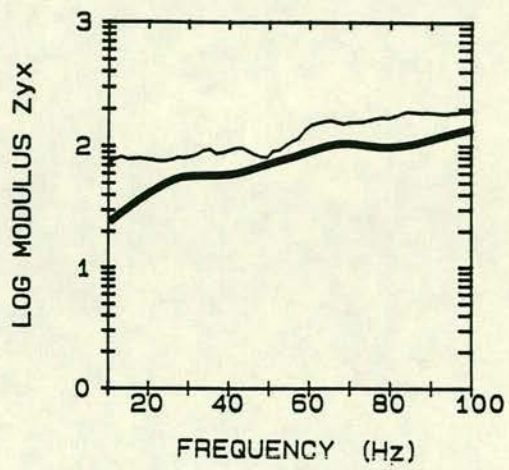
Z_{yx} Observational series



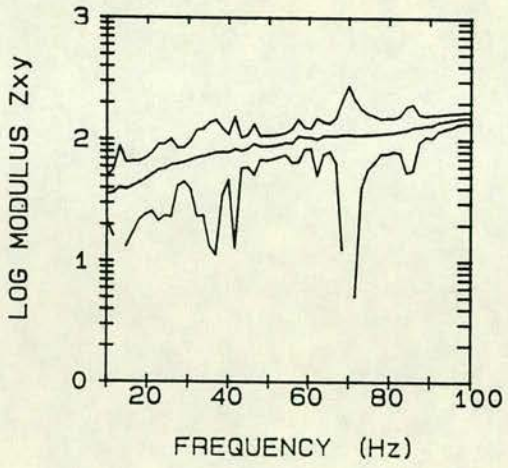
Z_{yy} Observational series

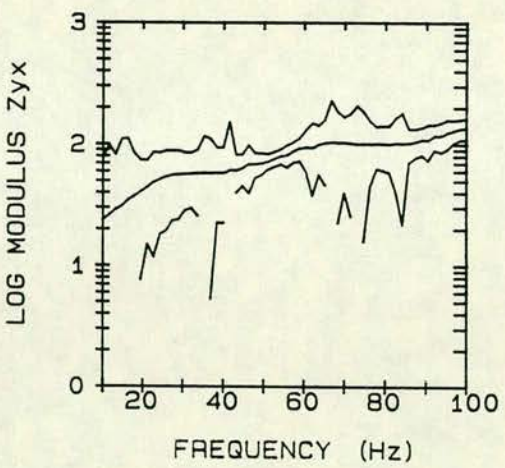
FIGURE 12 The frequency domain solutions (spectral impedance) of Example Site 1, based on MAXENT spectral analysis. Only the off-diagonal elements Z_{xy} and Z_{yx} are shown. **(a)** is a comparison between impedances with the WA method (thick solid line) and the spectral stacking method (thinner line). Error estimates are not included. **(b)** shows the impedance functions computed with the WA method; the error curves refer to one standard deviation. **(c)** shows the impedance functions computed with the WA method *after* data selection with the robust automatic scheme (A).


LOG MODULUS Zxy
3
2
1
0
20
40
60
80
100
FREQUENCY (Hz)

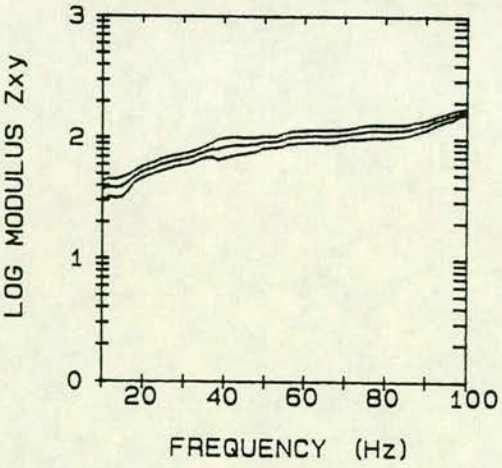

LOG MODULUS Zyx
3
2
1
0
20
40
60
80
100
FREQUENCY (Hz)

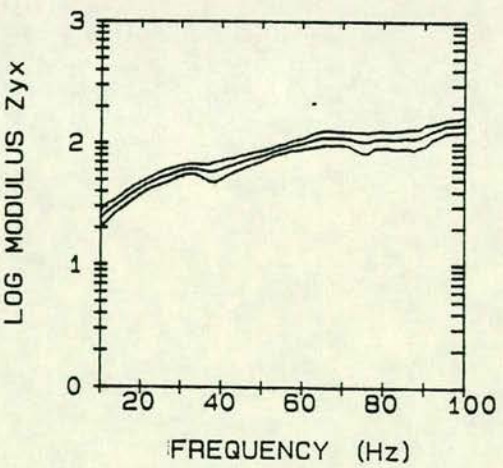
(a)


LOG MODULUS Zxy
3
2
1
0
20
40
60
80
100
FREQUENCY (Hz)


LOG MODULUS Zyx
3
2
1
0
20
40
60
80
100
FREQUENCY (Hz)

(b)


LOG MODULUS Zxy
3
2
1
0
20
40
60
80
100
FREQUENCY (Hz)


LOG MODULUS Zyx
3
2
1
0
20
40
60
80
100
FREQUENCY (Hz)

(c)

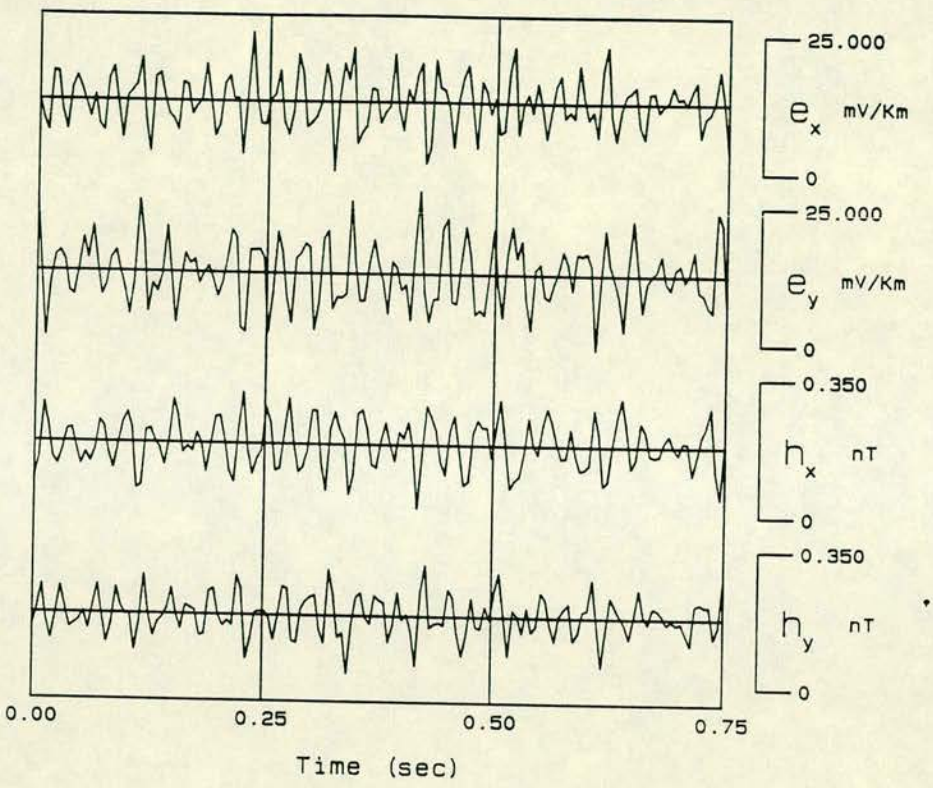


FIGURE 13 One 'outlying' data window, rejected by the automatic data selection scheme, and showing severe contamination by a 42Hz harmonic.

present an 'outlying' data window. Actually, this was 'selected' in an attempt to study the populations that tail off to the right and left of the $z_{xx}(0)$ and $z_{xy}(0)$ element histograms. It turns out that these populations consist of such data at various S/N ratios, in addition to weak transient waveforms. The dominant frequency here is a broad 42Hz harmonic of unknown origin. Note its effects on the stacked function at the same frequency range (Fig 12a); observe also that the WA result (Fig. 12c) is also slightly affected. This simple example is a case of a multimodal distribution forced into the data sample space by a repeating noise source. If one is prepared to spend some more time interpreting the histograms, Clark (1977), and Clark (1977), provide methods that will disect these distributions, thus facilitating quantitative measures of their statistics, and a more detailed perception of the sample space.

VIII.2 Example Site 2

The z(0) factor histograms for this site are given in Figure 14(a)-(d); they contain determinations from 495 data windows with N=50. It can easily be seen that the distributions of the z_{ij} populations are much more even at this site, than those of Figure 4; the inadequate noise sources are not as powerful in this site, as they have been for Example Site 1. The combination of inadequate and inconsistent noise sources, cause the spread observed in the histograms. However, the distributions are far from being normal; they appear to be skewed for the z_{yx}, and heavily tailed for the z_{yy} elements. Again, it can be shown that the WA procedure provides better results than spectral stacking, but such explicit comparisons will be omitted for brevity. The WA

(a)

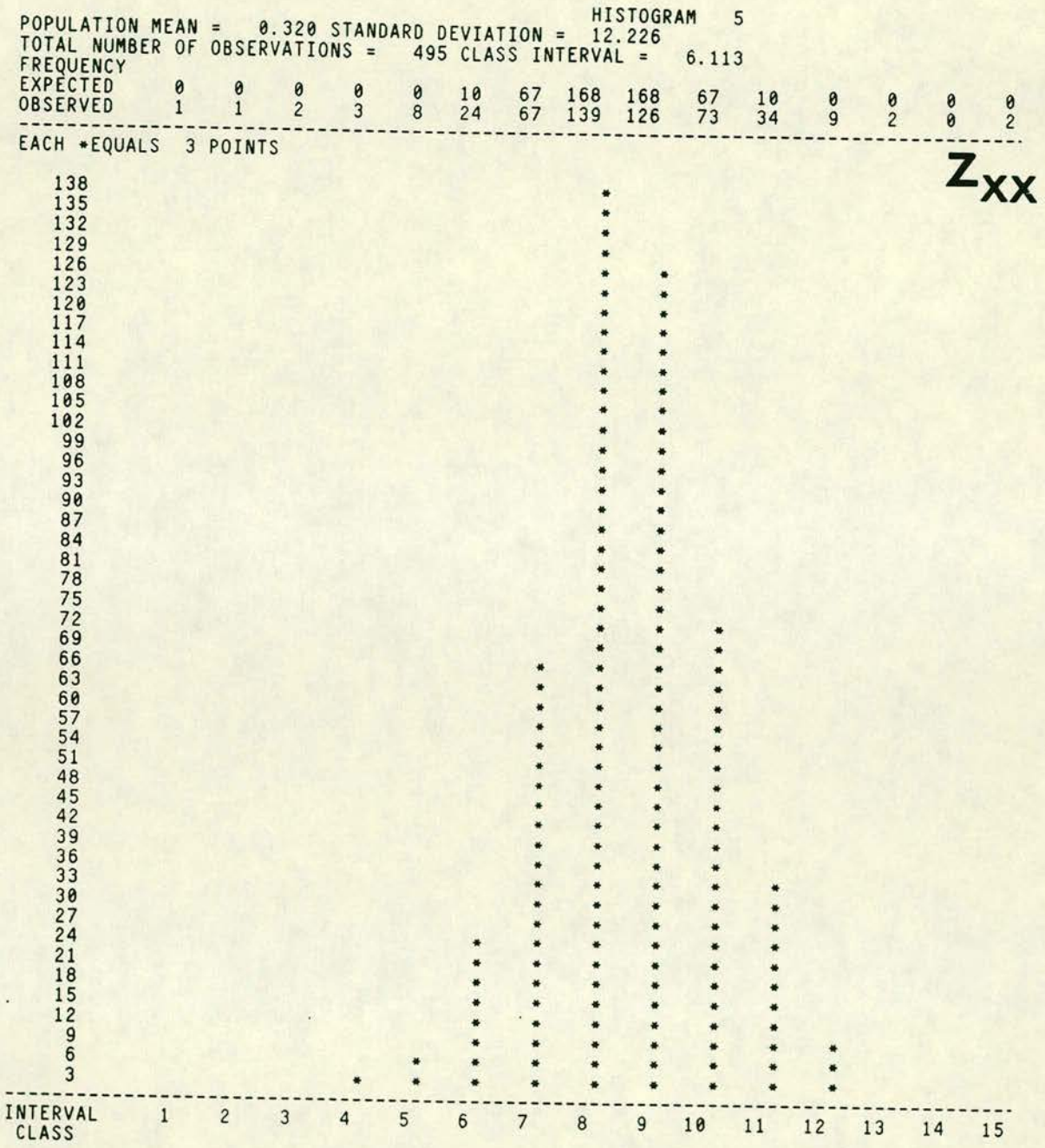


FIGURE 14 **(a)-(d)** The z_{ij} factor histograms for Example Site 2, as per Figure 8(a-d)

(b)

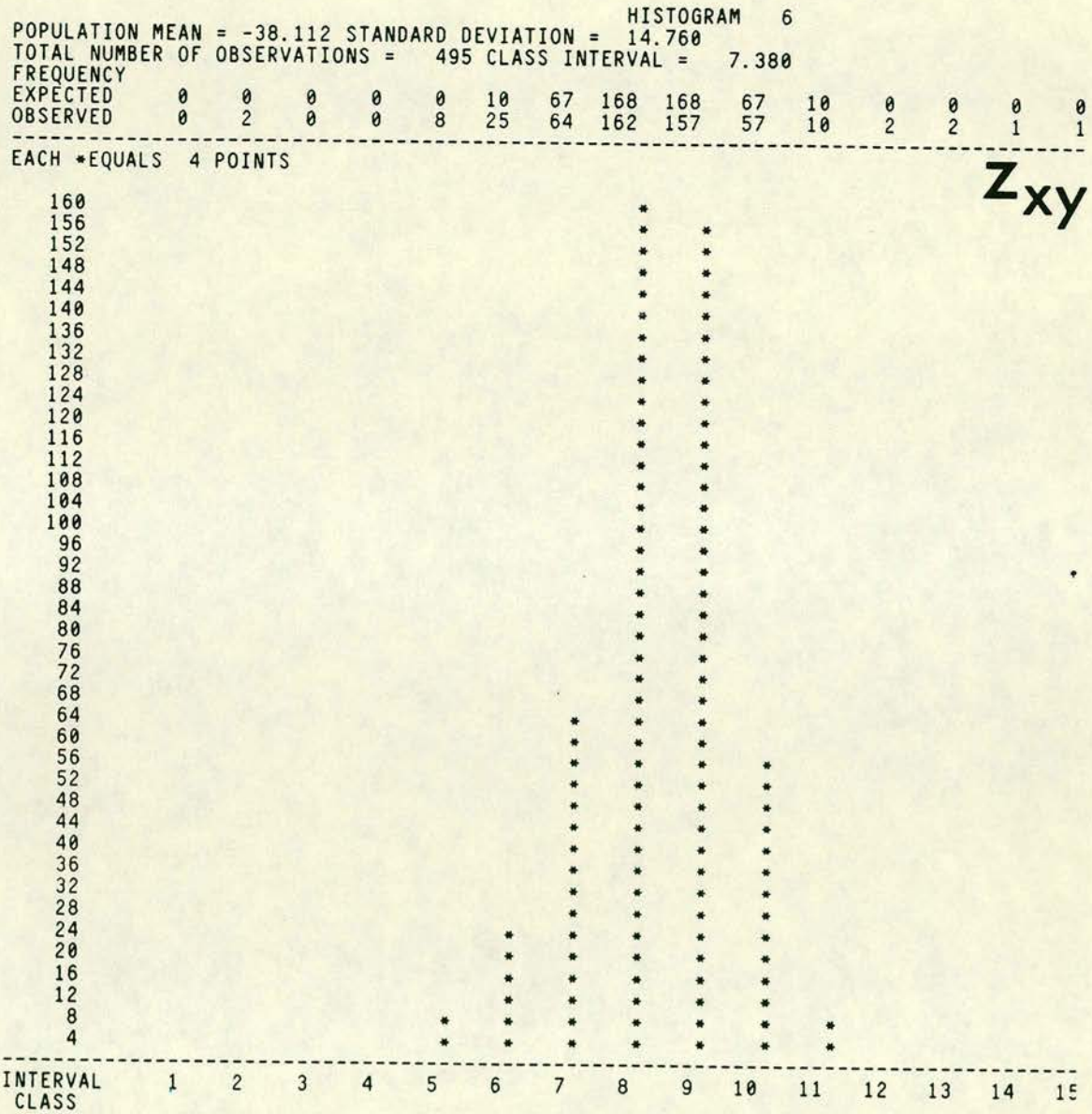


FIGURE 14 ..continued

(c)

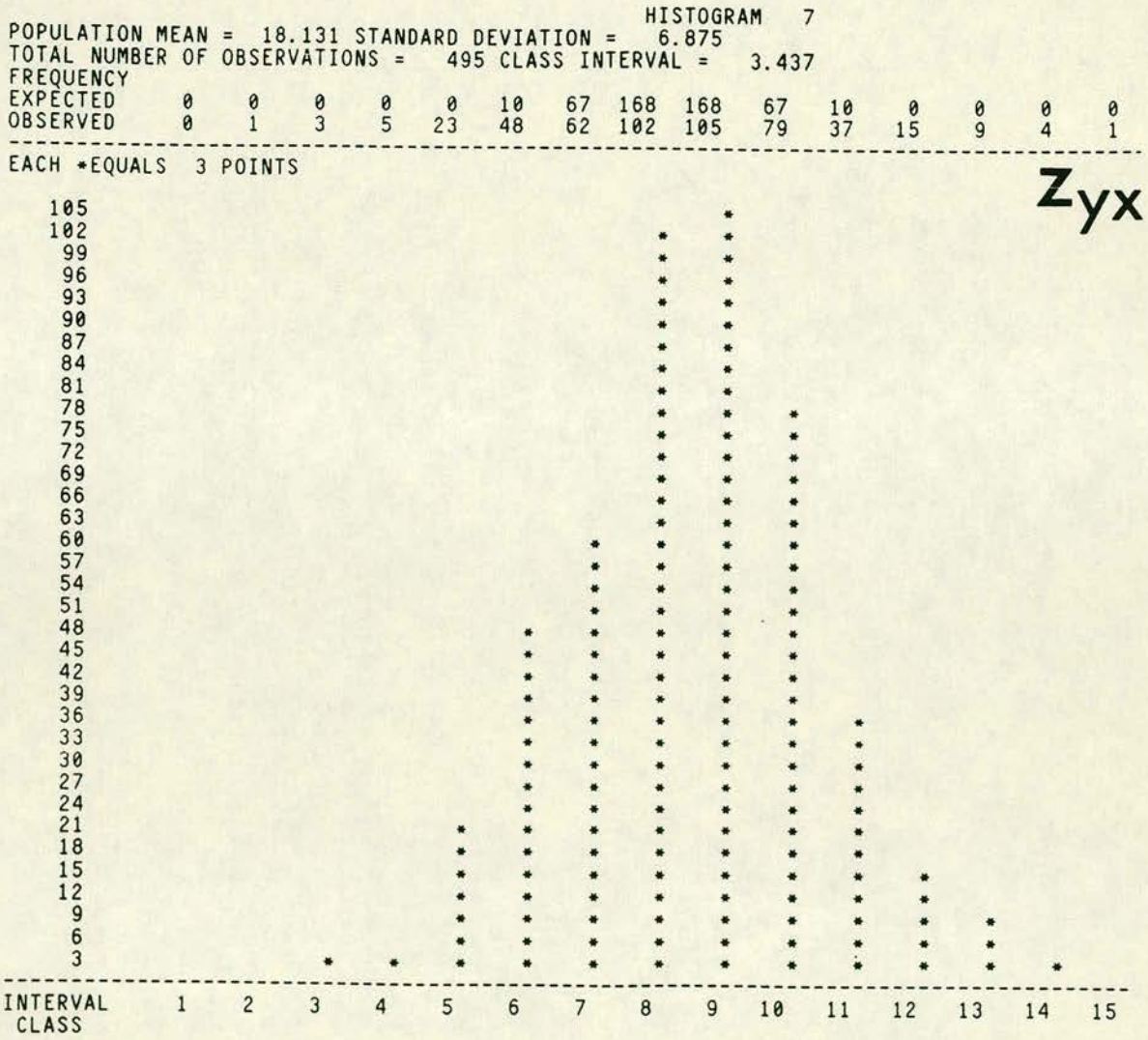


FIGURE 14 ..continued

(d)

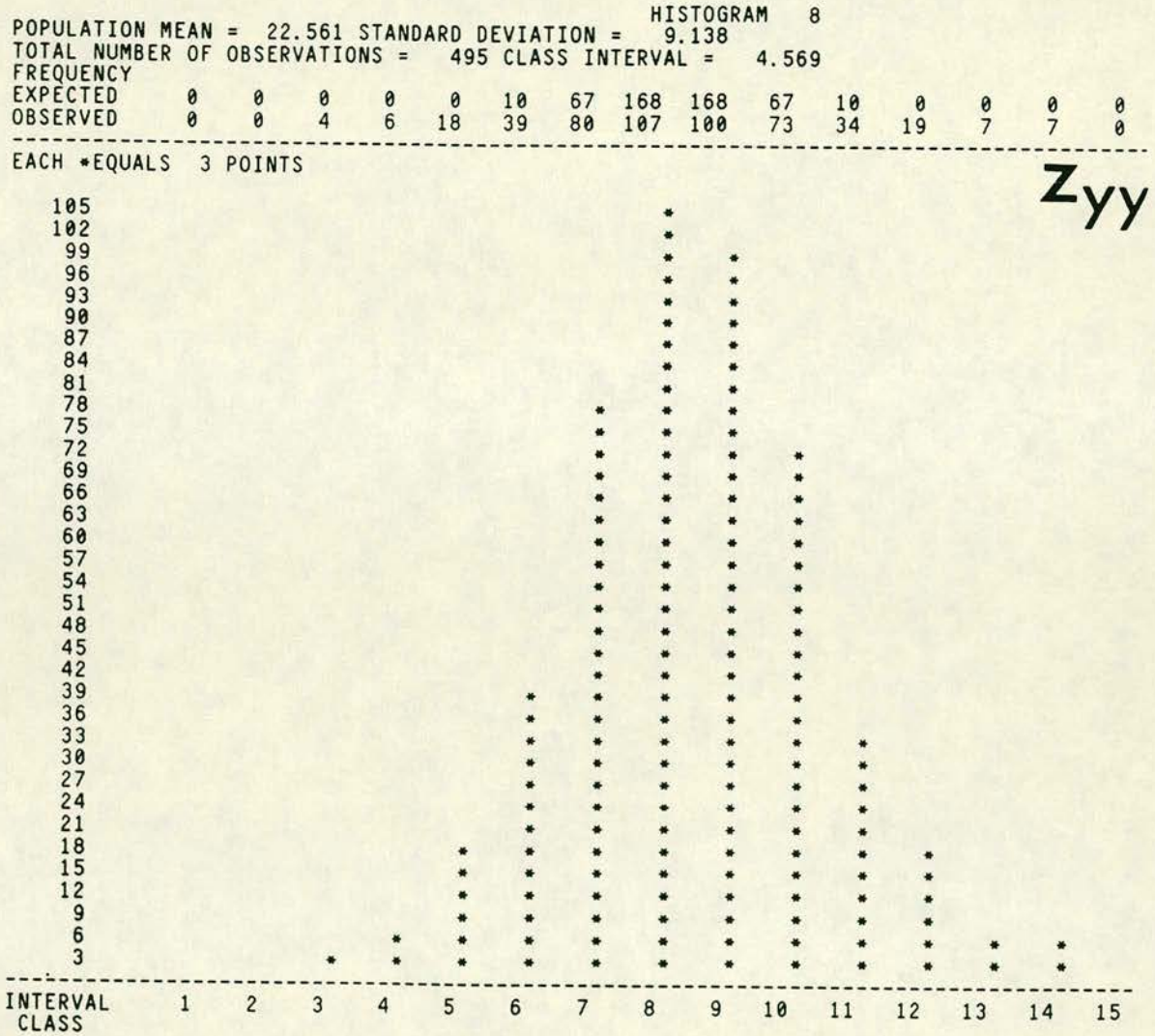


FIGURE 14 ..continued

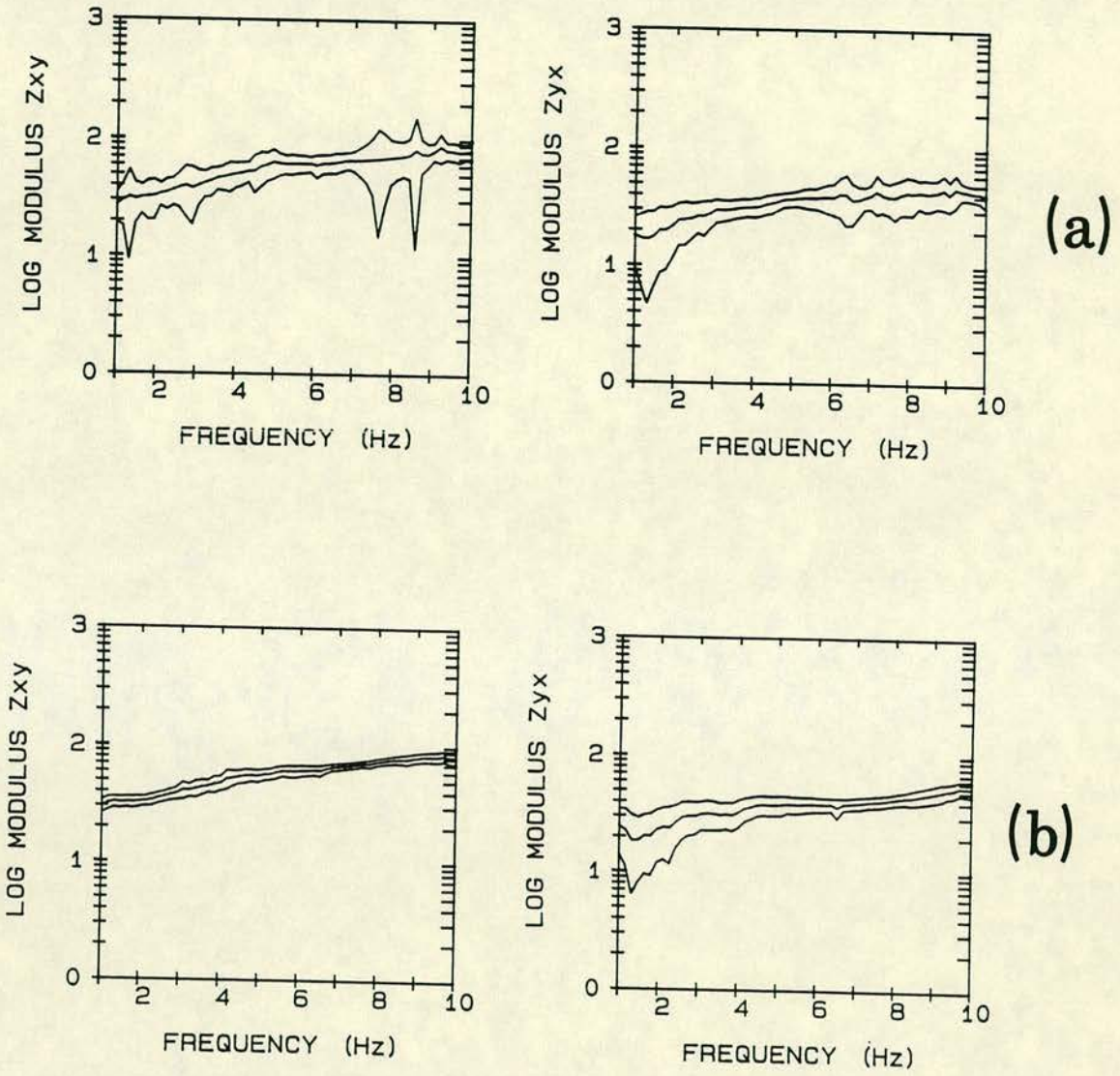


FIGURE 15 The frequency domain solutions (spectral impedance) of Example Site 2, based on MAXENT spectral analysis. Only the off-diagonal elements Z_{xy} and Z_{yx} are shown. **(a)** are the impedance functions computed with the WA method; the error curves refer to one standard deviation. **(b)** shows the impedance functions computed with the WA method after data selection with the robust automatic scheme (A).

impedance and its standard deviation are shown in Figure 15a, computed subject to the same coherence tests as per Example Site 1. Time domain data selection was applied in the same way as per Example Site 1. The frequency domain WA result of the 103 accepted data windows, is shown in Figure 15b. A considerable reduction of scatter can be observed again. Unfortunately, the obvious low frequency (<2.5Hz) noise problem is still unsolved; this is a pervasive feature of the data at this site, and little can be done about it, at least at the data selection level. This defines to some extent the limitations of the technique. Robust Frequency Domain procedures will reduce the scatter but will not necessarilly smooth or alter the shape of the the impedance curve; this is determined by the spectral components contained in the data. If such spectral contaminants are always present, or persist, all simple estimation procedures will sense a false 'signal', and will be unable to tell the difference...

IX. Discussion

In the present study, I have attempted to combine an investigation of the time-local properties of data infested with intense extraneous noise sources, with the presentation of techniques intended to remove them.

The concept of short data length processing can readily be recognised as advantageous. The geoelectric structure of the Earth is time invariant (stationary) for the time scales of EM surveys. The EM field however is not a stationary process. This applies for both the long period (MT) range (e.g. Dubrovsky and Kramarenko, 1977), and, most certainly, the short period (AMT) range. Beamish and Tzanis (1986) and Tzanis and

Beamish (1987a) have shown that that the sub-audio and audio frequency fields can be approximated with locally stationary processes at best. Moreover, noise is, in general, non-stationary; not only anthropogenic noise sources, such as the cases presented, but many forms of naturally induced perturbations as well (e.g. wind noise). Its statistical behaviour becomes more important as its intensity increases. A logical consequence of these observations would be try to capitalize on this short term EM field behaviour by adapting to it and analysing short data lengths, while, at the same time subjecting it to strict acceptance tests. The conventional spectral analysis technique (FFT) however, suffers from severe statistical and resolution limitations when the data length becomes too short (e.g. less than N=50). An FFT based analysis cannot cope with such lengths, especially when they are discontinuously recorded. Thus, scrutiny into the time-local properties of field data is difficult at best. In fact, it is conveniently forgotten in view of the problems that spectral estimation presents. The introduction of the real time cascade decimation spectral analysis method (Wight et al, 1977), with its short data length handling capacity, is partly due to the recognition of these problems and an important step towards exploiting the time-local properties of the field data. The introduction of MAXENT spectral analysis herein, is an additional attempt for a high resolution tool that will allow scrutiny into both the time-local and frequency-local properties of the field data. The examples presented in this study are quite illustrative, and indicate, in my opinion, that the subject requires further and serious consideration.

The data selection techniques introduced here, will work best when used in association with short data length processing. Naturally, they can

be used to process data of *any* length. However, the time-local variation of noise content will be more easily delineated if it is not 'obscured' by the preceding or following realisations of the EM field. One must bear in mind that the formation of the time domain covariances is an extremely effective averaging operation. The techniques are shown to be quite effective when applied to data contaminated by inadequate noise sources. But they cannot be a panacea. Proper use requires some conditions to be upheld by the data. These can be listed as follows.

(a) Because of the aforementioned averaging properties of covariance formation, the techniques are very sensitive to relatively intense inconsistencies/inadequacies in the data. Physically, this is a consequence of using total energy conservation considerations only; it means that the technique will perform better when the noise terms are of the same order of magnitude as the natural EM field. Weak noise sources will tend to be 'smeared out' and will stand out less clearly than in the example sites 1 and 2. This property must be taken into account when interpreting the histograms.

(b) The above reasoning suggests that the behaviour of the **z**(0) matrix strongly depends on the existence of dominant waveforms in the data. Long period data, possess such properties ('red' spectra) by definition. Therefore, the contribution of the near Nyquist components will be disproportionately low, when compared to those near the fundamental; their influence will be overshadowed by the power of the relatively long periods when forming the covariances, and, the **z**(0) matrix will tend to reflect the linearity conditions of the longer period part of the spectrum. It follows that the techniques are suitable for data possessing (nearly) white spectra, e.g. data in the sub-audio and

audio-frequencies (> 0.1Hz). Otherwise the data will have to be prewhitened for a proper implementation.

(c) Implementation of Technique (B) requires a somewhat extensive sample space; it is a statistical approach. This may be difficult if long data windows (e.g. N=256 or 512) are used, due to possible storage limitations. I believe that this can serve as an extra argument for its use with short data lengths, since in this way we can easily afford a large sample space. Moreover, it cannot be used in real time, although all the necessary parameters can be accumulated and stored during data collection. Only Technique A is a very fast criterion that can be used for real time data quality evaluation.

The effectiveness of the techniques has been demonstrated by the above examples. The following *advantages* can be enumerated.

(a) They are very simple, to program and operate. They do not require any special mathematics for their implementation.

(b) They do not require large storage capacity and can be used in very small computer systems. In this respect such a scheme is also very fast. The automatic selection scheme of Technique B for instance, requires about 7 minutes, including input/output operations from a hard disk, to process 495 data windows on a PDP-11/73 system.

(c) Finally, Technique B is very versatile. It permits the selection of *any* combination of data populations (manual scheme B) for further study and scrutiny of their properties.

I will not devote much more space to discuss the advantages and merits of robust estimation. Its performance and value is made perfectly clear

by the examples presented. I believe that robust procedures can offer significant advances in the sensitive field of (A)MT parameter estimation, where small changes in the location of the estimates can lead to large deviations in the parameters of the geoelectrical models that 'generate' the observations. However, a note of caution applies with respect to which robust scheme is to be chosen. As usually happens with parameters requiring statistical estimation, there exist no general methods or 'canned' procedures that solve all problems ex machina. I have tried to show this by demonstrating how simple variants of the same basic estimation procedure produce wildly different results (i.e. RLS/RWLS), and, even when the better procedure (RWLS) was chosen, how heavily it relied on the numerical method employed, in order to solve the system (Normal Equations vs Golub). Eventually, I ended up implementing a procedure that is both general and stable, and has been proven using heavily contaminated AMT data. But even for this procedure, the same note of caution applies: It must not be used as a 'black box' in neither the time, or frequency domains, because it is designed to improve the chances of correct parameter location, without offering assurances. This defines the degree of caution that is required when estimating the parameters of an ill-posed problem, such as the (A)MT problem for many types of data.

APPENDIX 4.A Derivation of random error formulae fot the 1-input/ 1-output time domain factor z(0).

As stated in the main text this appendix serves only as a demonstration of the statistical isomorphism between the first order time domain, and frequency domain processes; the method that will be followed here derives from that of Bendat and Piersol (1971). Let us begin by stating the variables of our problem. Let

$$\varepsilon(t) = e(t) - \sum_{\tau} z(\tau)\, h(t-\tau) \qquad \text{(A1)}$$

be the true, and

$$\hat{\varepsilon}(t) = e(t) - \sum_{\tau} \hat{z}(\tau)\, h(t-\tau) \qquad \text{(A2)}$$

be the observed 1-input 1-output waveform shaping relationship between the electric vector e(t) and the magnetic vector h(t). By rewriting (A2) as

$$e(t) = \hat{\varepsilon}(t) - \sum_{\tau} \hat{z}(\tau)\, h(t-\tau) \qquad \text{(A3)}$$

letting $\tau=0$, and substituting into (A1) we obtain our observational equation

$$\varepsilon(t) = \hat{\varepsilon}(t) - [\hat{z}(0) - z(0)]\, h(t) \qquad \text{(A4)}$$

Let $\hat{z}(0) = \hat{z}$, and $z(0) = z$, and denote the auto- and cross covariance functions by σ^2. The total energy in the error series, is given by

$$\sum_t \varepsilon(t)\varepsilon(t) = \sum_t \hat{\varepsilon}(t)\hat{\varepsilon}(t) + 2 \sum_t \hat{\varepsilon}(t)\,(\hat{z}-z)\, h(t) + \sum_t (\hat{z}-z)^2 h^2(t)$$

$$= \hat{\sigma}^2_{\varepsilon\varepsilon} + \sigma^2_{hh}(\hat{z}-z)^2 + 2\left[\sum_t e(t)\, h(t)\,(\hat{z}-z) - \sum_t \hat{z}\, h(t)\,(\hat{z}-z)\right]$$

$$= \hat{\sigma}^2_{\varepsilon\varepsilon} + \sigma^2_{hh}(\hat{z}-z)^2 + 2\left[(\sigma^2_{eh} - \hat{z}\,\sigma^2_{hh})(\hat{z}-z)^2\right] = \sigma^2_{\varepsilon\varepsilon}$$

However, since $\sigma^2_{eh} = \hat{z}(0)\,\sigma^2_{hh}$, the cross term in the above expression vanishes so that

$$\sigma^2_{\varepsilon\varepsilon} = \hat{\sigma}^2_{\varepsilon\varepsilon} + \sigma^2_{hh}|\hat{z} - z|^2 \qquad \text{(A5)}$$

The variances $\sigma^2_{\varepsilon\varepsilon}$, $\hat{\sigma}^2_{\varepsilon\varepsilon}$ and σ^2_{hh} are variances of normally distributed quantities, therefore they possess χ^2 distribution properties. If the

variances are computed from a time series of length N, then, $\sigma^2_{\varepsilon\varepsilon}$ has n=N degrees of freedom assuming that it belongs to a zero-mean population, or, N-1 degrees of freedom if we assume that estimation of the mean imposes one constraint. The terms of (A5) are all statistically independent, and therefore the following relationship should hold

$$\chi^2 / n = \chi^2_{n_1} / n + \chi^2_{n-n_1} / n \qquad \text{(A6)}$$

From (A2) one can see that $\hat{\sigma}^2_{\varepsilon\varepsilon}$ is defined in terms of $\hat{z}$, which constitutes one independent constraint, so that

$$n \, \hat{\sigma}^2_{\varepsilon\varepsilon} / \sigma^2_{\varepsilon\varepsilon} = \chi^2_{n-1} \qquad \text{(A7)}$$

and as a consequence,

$$(n \, \sigma^2_{hh} \, |\hat{z} - z|^2) / \sigma^2_{\varepsilon\varepsilon} = \chi^2_1 \qquad \text{(A8)}$$

has only one degree of freedom. From (A2), it follows easily that

$$\hat{\sigma}^2_{\varepsilon\varepsilon} = \sigma^2_{ee} - |\sigma^2_{eh}| / \sigma^2_{hh} \Rightarrow \hat{\sigma}^2_{\varepsilon\varepsilon} = \sigma^2_{ee}[1 - R^2_{eh}]$$

where R^2_{eh} is the correlation coefficient between e(t) and h(t), given by $R^2_{eh} = |\sigma^2_{eh}| / \sigma^2_{ee} \, \sigma^2_{hh}$. Consequently, (A7) gives

$$n \, \sigma^2_{ee}[1 - R^2_{eh}] / \sigma^2_{\varepsilon\varepsilon} = \chi^2_{n-1} \qquad \text{(A9)}$$

Combine (A8) and (A9) in terms of an F distribution with $n_1 = n-1$, and $n_2 = 1$ degrees of freedom, to obtain

$$F_{1,n-1} = \frac{(n-1) \, \sigma^2_{hh} \, |\hat{z} - z|^2}{\sigma^2_{ee}[1 - R^2_{eh}]}$$

By defining the upper limit of the random error such that $|\hat{z} - z|^2 \leq \varepsilon^2_r$, we get the final expression

$$\varepsilon^2_r = \frac{1}{n-1} F_{1,n-1:a} [1 - R^2_{eh}] \frac{\sigma^2_{ee}}{\sigma^2_{hh}}$$

The expressions for the random error in the 2-input 1-output case can be obtained by extrapolation of these results. The reasoning remains the same, but the statistics become much more complicated. The frequency domain analysis that can be used as a guide, can be found in Goodman (1965).

APPENDIX 4.B Random Errors and Multiple Coherent Noise Waveforms.

This brief study will be carried out entirely in the Frequency Domain, and will use the formulation of Pedersen (1982) for the random error of the MT impedance function, i.e.

$$E_r^2(\omega) = \frac{1}{n-4} F_{1,n-4:a} \frac{[1 - \gamma^2_{i.hd}]\ (E_i E_i^*)}{[1 - \gamma^2_{hd}]\ (H_j H_j^*)}\ ,\ i=x,y,\quad j=h,d$$

where n is the number of degrees of freedom associated with the measured spectra, and F is the 100a percentage point of an F distribution with n-4 and 1 degrees of freedom. In accordance with the convention of using lower case symbolization for the time domain and upper case for frequency domain processes, I use the symbol $E(\omega)$ to denote the 68% confidence limit, based on the above expression for the variance, i.e.

$$E(\omega) = \sqrt{E_r^2(\omega)} \ / \ 2.$$

In the following discussion, the quantity $E(\omega)$, or its square will be used throughout. Notably, for the same controlling parameters, Pedersen's formulation produces about half the magnitude for the random error, with respect to that of Goodman (1965).

There are very sparse references to the properties and behaviour of the random errors, if any at all. If they are mentioned, this usually happens on a casual basis, while a systematic description of their usefulness doesn't exist. The author is only aware of the original theoretical study of Pedersen (1982), which provided no practical examples. Also, a theoretically related, but quite different study by Pedersen and Svennekjaer (1984) can be used as a reference. Also Beamish (1986), reports their implementation as weights, in estimating the elements of the MT impedance tensor. The present study is an empirical approach, aiming to demonstrate that these overlooked

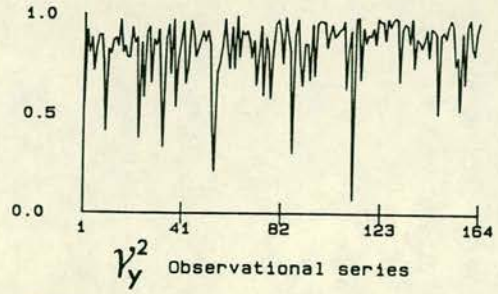
quantities are potentially useful indicators of data quality, and also, potential weight factors, for the Frequency Domain estimation of the impedance tensor. I shall make use of data from a measurement site with noise characteristics similar to these of Example Site 1, (i.e. infested by powerful transient events), but displaying considerably more coherent frequency domain solutions.

Figure B1 shows the observational series of $|Z_{yx}|$ and $|Z_{yy}|$, and their associated errors E_{yx} and E_{yy}, for two frequencies, of 25Hz and 35Hz respectively. The series has been produced with a limit of $\gamma_y^2 = 0$, i.e. the response of all recorded data realizations is displayed, at these frequencies. The $|Z_{yx}|$ and $|Z_{yy}|$ series display two types of solutions. The spikes, that correspond to high energy transient noise waveforms, and the low amplitude 'data background' that displays a random behaviour. Scrutiny reveals that this 'random' behaviour is closely related to the behaviour of the predicted coherence, and, in general, will display larger errors for less coherent data. Thus, the figure demonstrates what is clearly expected, i.e. large errors associated with incoherent data realizations, and also what is quite unexpected, i.e. that the largest error spikes are not necessarily associated with low coherence and/or 'background' data structure. As a matter of fact, some of them are associated with quite high coherence values. This behaviour is very similar to that of the time domain solutions for the Example Site 1.

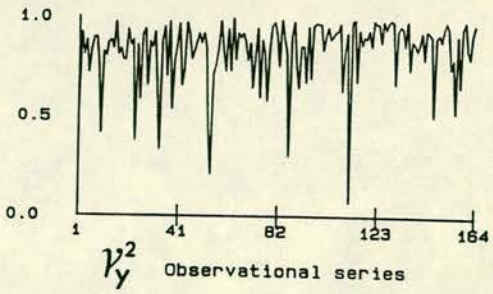
Figure B2 helps to understand this behaviour by displaying the same data series, but this time selected so that $\gamma_y^2 > 0.9$, for every impedance element determination shown. This figure is quite revealing. It becomes immediately clear that perfectly coherent data display great

deviations from the expectation values, and huge errors. All the spikes in the $|Z_{yx}|$ and $|Z_{yy}|$ series can be paired with the spikes that pop out of otherwise consistent random error populations. Moreover, all the spikes observed in the impedance function can be traced down to powerful noise waveforms. The consistency of the 'data background' should be expected, since it is a result of essentially coherent adequate data. The spikes in the $|Z_{yx}|$ and $|Z_{yy}|$ observational series are also understandable, in terms of the powerful transient events generating them. The behaviour of the random error however, is much more complicated. Figure B3 contains the observational series of all the factors that control the magnitude of $E(\omega)$, for a given number of degrees of freedom, and F-distribution percentage point. The plots display, from bottom to top, (1) the magnitude of the 68% confidence limit, (2) the absolute power of the observed E-field, (3) the absolute power of the observed H-field, (4) the input (magnetic) field coherence, and (5) the predicted coherence function. We observe that as a rule, large random errors are associated with high power E and/or H fields, of inconsistent nature, i.e. erratic values of the ratio $(E_y E_y^*)/(H_x H_x^*)$. In many cases, large random errors are associated with high values of the input field coherence. Both features are almost textbook cases of extremal bias (also see Pedersen and Svennekjaer, 1984). However, in cases of inconsistent multiple coherent noise populations such as the ones under scrutiny, no rule whatsoever can be obtained from the study of Figure B3, or any related feature, and no pattern emerges as to how the random error behaves with respect to the controlling parameters (at least I haven't been able to extract any). To the best of my understanding, the factors controlling the random error behave randomly. Therefore, one cannot say that a single powerful telluric noise pulse

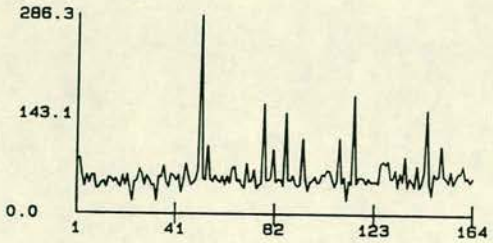
FIGURE B1 The frequency domain observational series of the $|Z_{yx}|$ and $|Z_{yy}|$ impedance tensor elements, their associated E_{yx} and E_{yy} Pedersen (1982) random errors, and the predicted coherence $\gamma^2_{i.hd}$. Single frequency parameters for **(a)** 25Hz, and **(b)** 35Hz.



γ_y^2 Observational series

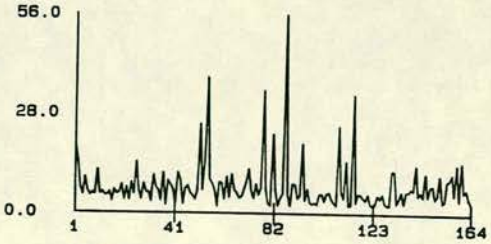


γ_y^2 Observational series

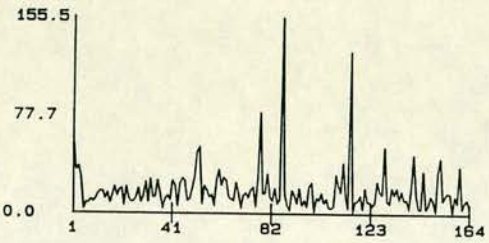


Z_{yx} Observational series

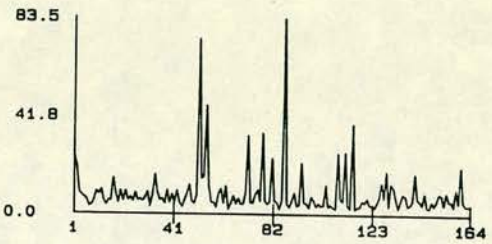
F = 25Hz



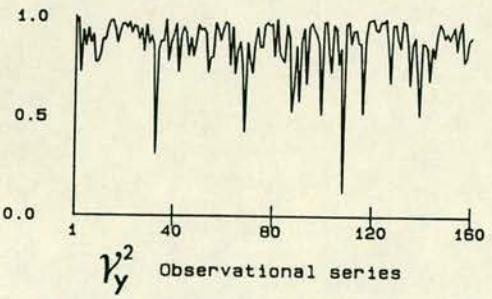
E_{yx} Observational series



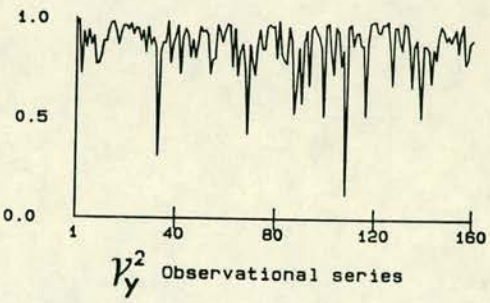
Z_{yy} Observational series



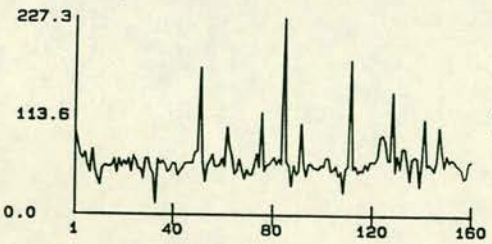
E_{yy} Observational series



γ_y^2 Observational series

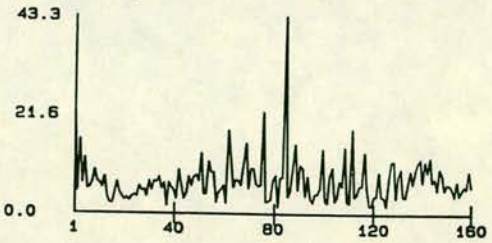


γ_y^2 Observational series

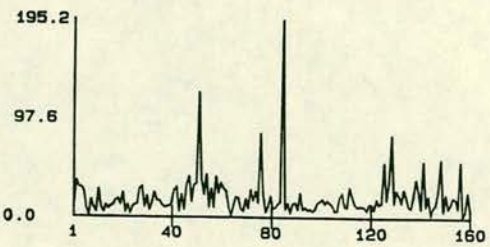


Z_{yx} Observational series

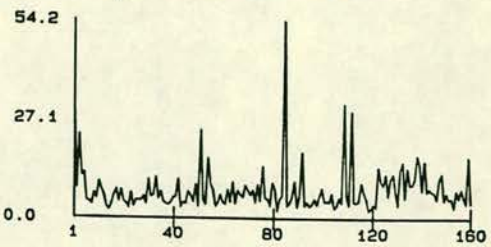
F = 35Hz



E_{yx} Observational series

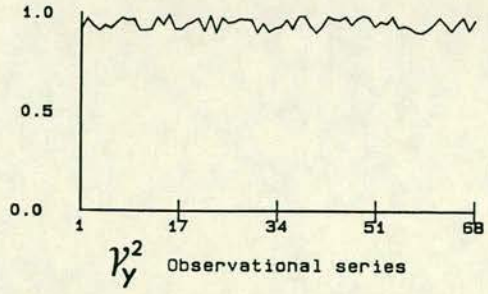


Z_{yy} Observational series

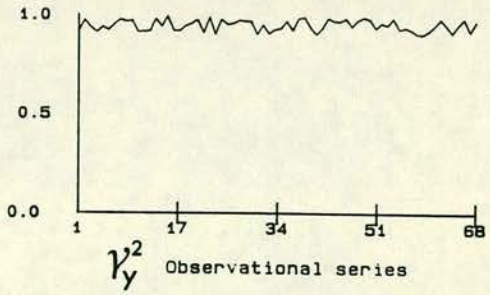


E_{yy} Observational series

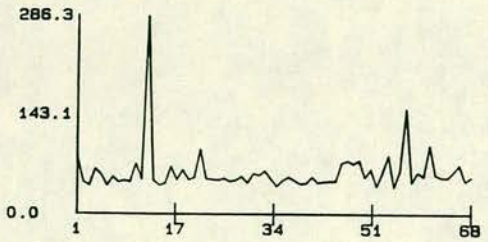
FIGURE B2 As per Figure B1, but only elements with predicted coherence $\gamma^2_{i.hd} > 0.9$ are shown.



γ_y^2 Observational series



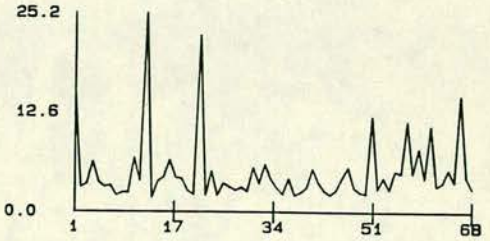
γ_y^2 Observational series



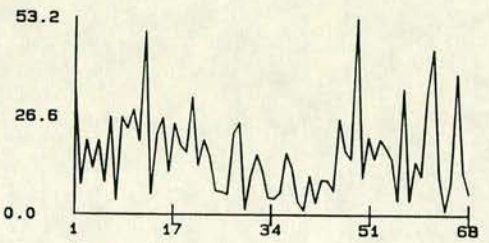
Z_{yx} Observational series

$\gamma_y^2 > 0.9$

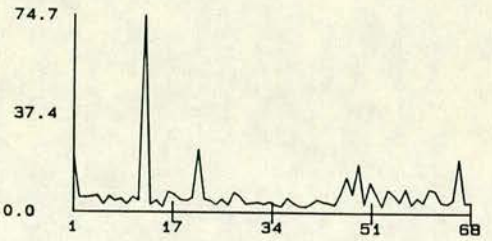
F = 25Hz



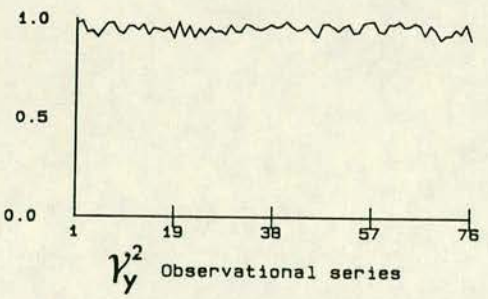
E_{yx} Observational series



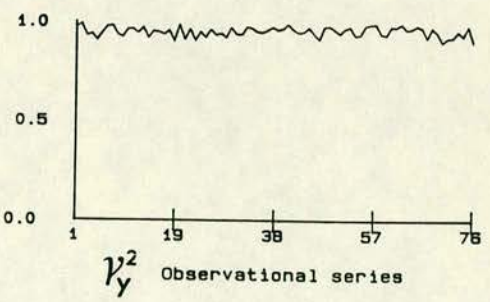
Z_{yy} Observational series



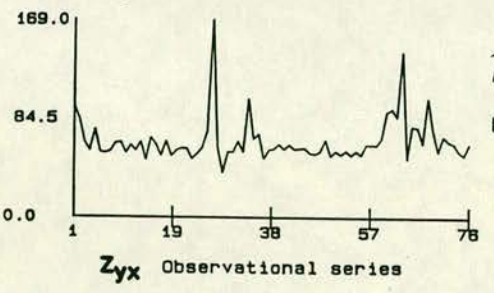
E_{yy} Observational series



γ_y^2 Observational series



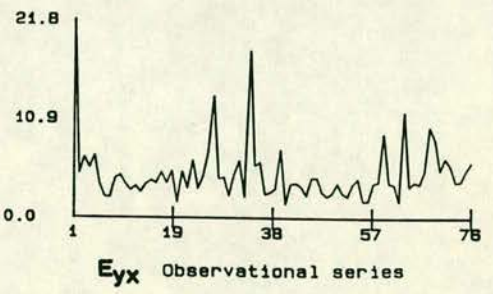
γ_y^2 Observational series



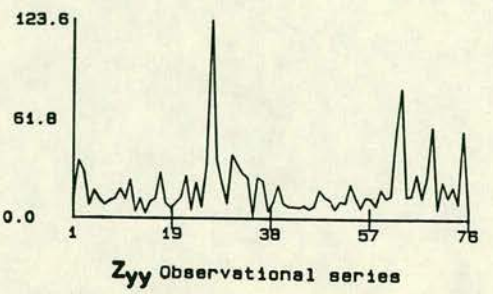
Z_{yx} Observational series

$\gamma_y^2 > 0.9$

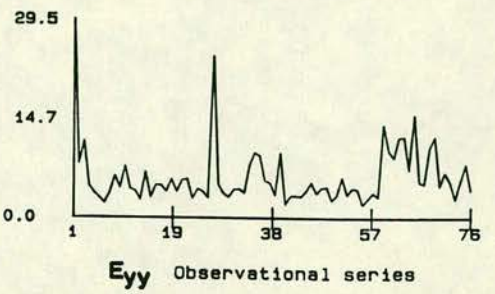
F = 35Hz



E_{yx} Observational series



Z_{yy} Observational series



E_{yy} Observational series

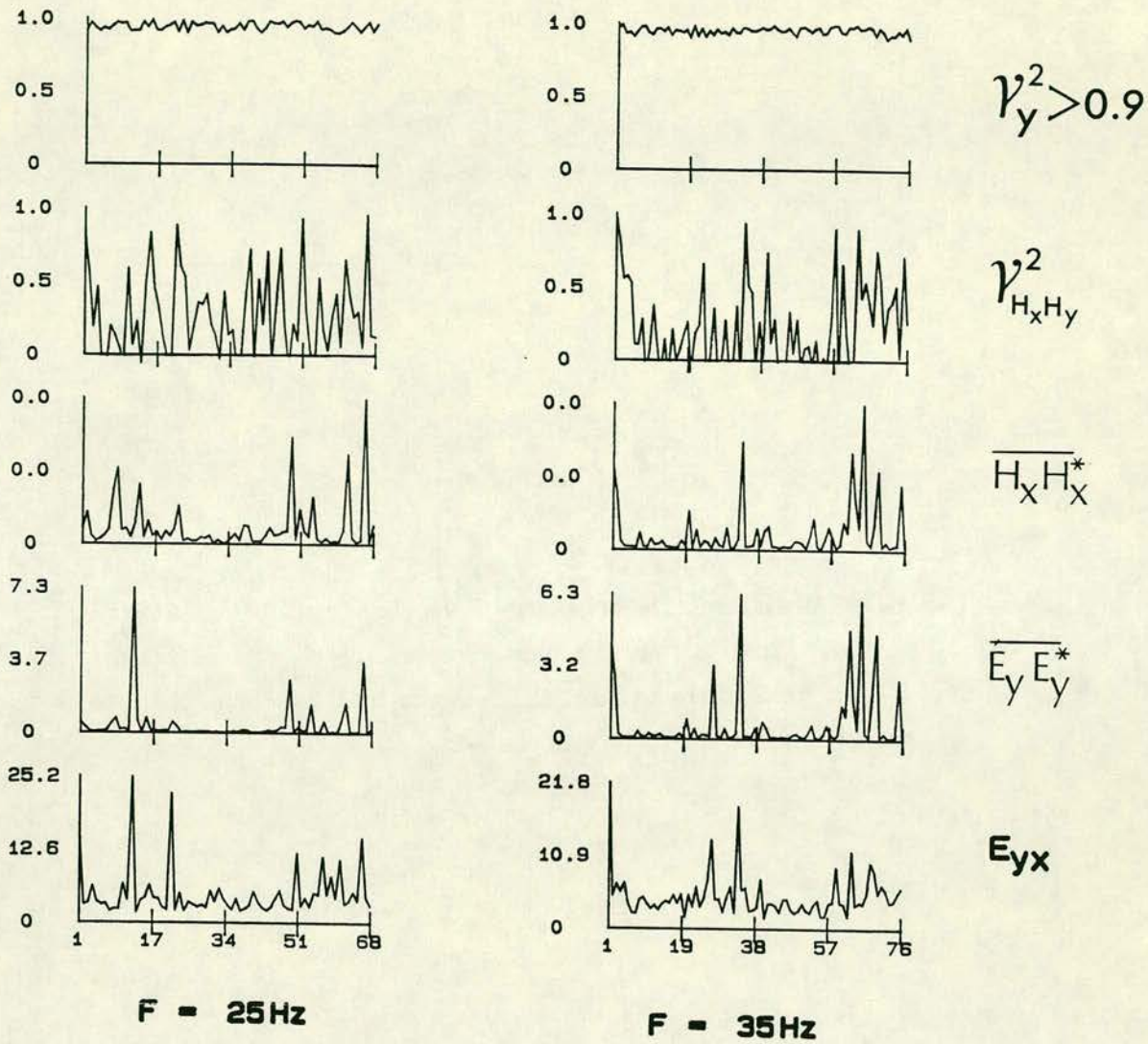
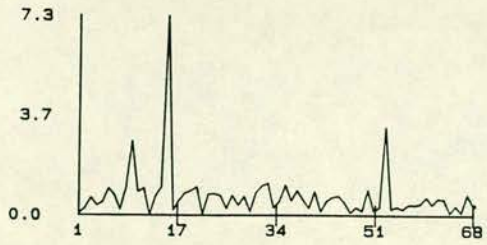
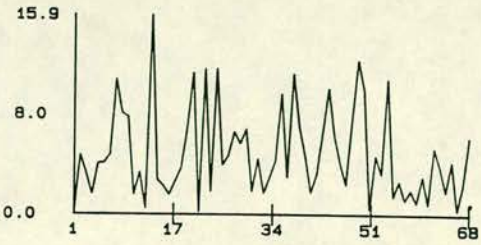


FIGURE B3 Simultaneous plot of the factors controlling the magnitude of the random error E_{yx}, for the observational series of Figure B2. These include, from bottom to top, the E_y telluric component auto-power, the H_x magnetic component auto-power, the input (magnetic) field ordinary coherence, and the multiple (predicted) coherence function. Overbars symbolize band averaging of FFT raw spectra.

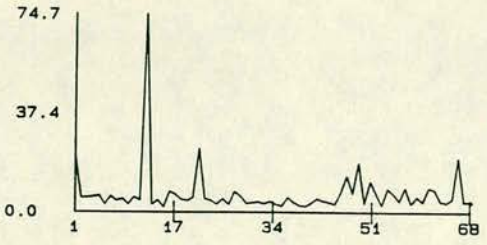
FIGURE B4 Random error weighting of the Figure B2 observational series. From bottom to top are depicted the $|Z_{ij}|$ observational series, the corresponding E_{ij} observational series, and the weighted series $|w\ Z_{ij}|$, with $w = 1/E_{ij}^2$.



$W.Z_{yy}$ Observational series



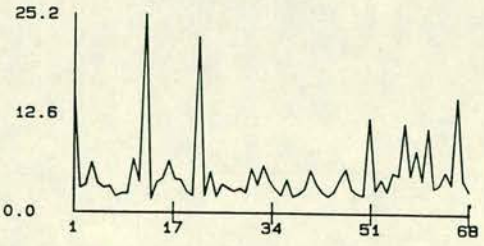
$W.Z_{yx}$ Observational series



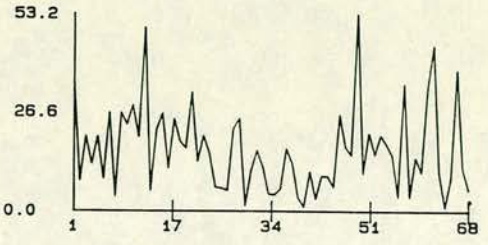
E_{yy} Observational series

$\gamma_y^2 > 0.9$

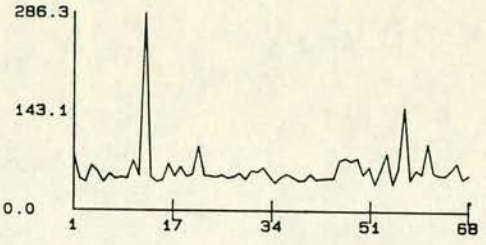
F = 25Hz



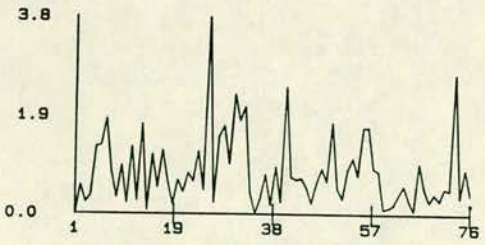
E_{yx} Observational series



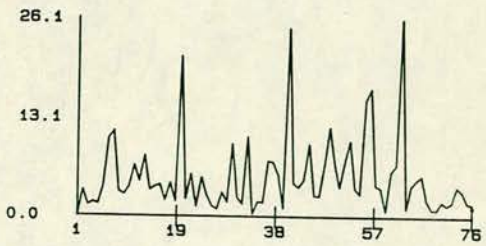
Z_{yy} Observational series



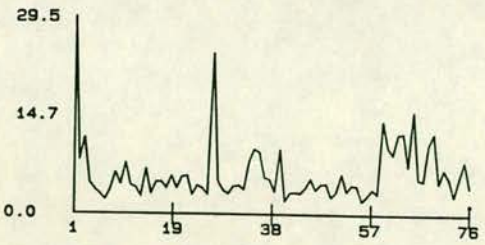
Z_{yx} Observational series



$W.Z_{yy}$ Observational series



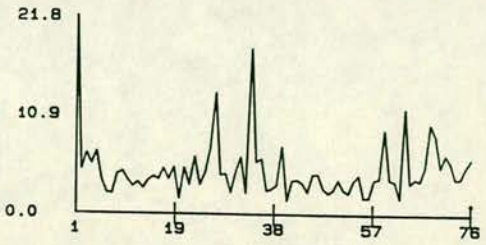
$W.Z_{yx}$ Observational series



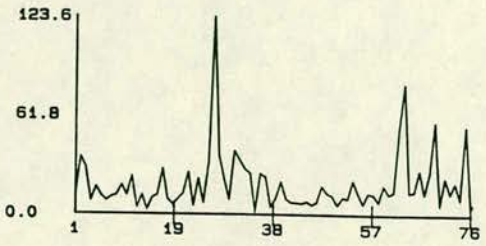
E_{yy} Observational series

$\gamma_y^2 > 0.9$

F = 35Hz



E_{yx} Observational series



Z_{yy} Observational series

Z_{yx} Observational series

will produce large errors, if the other factors do not contribute, and their interaction is, in general, non-linear. As stated in VII and restated here, there are no rules which produce produce guarantees about the random error behaviour; there exist only chances (or likelihoods). However, as can be seen in Figure B3, huge error spikes are always associated with quite bizarre field properties, and therefore a large random error emerges as an unequivocal indicator of data quality, even if perfect linear models can be fitted to the noisy data.

As such, the random error serves as:

(1) Measure of poor estimation due the incoherent behaviour of the EM field components controlling it, and

(2) Measure of poor estimation due to the bizare behaviour of the EM field components controlling it, although not always. Given however that we never have prior information about the behaviour of the EM field, the random error emerges as a potentially useful diagnostic aid and data processor. I have been using the random error in its second form, as a weight; this appears to be extremely useful in Frequency Domain processing. In order to avoid lengthy discussions, I introduce Figure B4, which displays the effect of random error weighting on the impedance series of Figure B2. The figure shows, from bottom to top, the $|Z_{ij}|$ observational series, the corresponding E_{ij} (error) series, and finally, the *weighted* observational series, denoted as $w.Z_{ij}$, where $w = 1/E_{ij}^2$. It can easily be seen, that the weighted observational series contains several spikes. These spikes however, correspond to the $|Z_{ij}|$ determinations displaying the *smallest* random errors; in addition, these determinations always belong to the 'data background'. The noisy $|Z_{ij}|$ spikes are larglly reduced, or even totally anihilated. These figures produce a clear example of how it is possible to obtain

reliable and useful ensemble averages, with the simplest means possible, even in the dreaded case of multiple coherent noise. The application of a robust procedure here, would only serve to slightly 'adjust' the WA estimate to the optimal location.

REFERENCES

Adam A. and Veró J., 1976, Magnetotelluric data processing methods, in KAPG Geophys. Monogr. Geoelectric and Geothermal Studies, pp.256-263, Adam, A. (editor), Akademiai Kiado, Budapest.

Arfken G., 1985, Mathematical Methods for Physicists (3rd edition), Academic Press, New York.

Bahr K., 1983, Joint interpretation of Magnetotelluric and Geomagnetic Data and Local Telluric Distortions, J. Geomag. Geoelectr. **35**, 555-566.

Banks R.J., Beamish D. and Geake M.J., 1983, Magnetic variation anomalies in northern England and southern Scotland, Nature **303**, No.5917, 516-518.

Banks R.J. and Beamish D., 1984, Local and regional induction in the British Isles, Geophys. J. R. astr. Soc. **79**, 539-553.

Beamish D. and Banks R.J., 1983, Geomagnetic variation anomalies in nothern England: processing and presenatation of data from a non-simultaneous array, Geophys. J. R. astr. Soc. **75**, 513-539.

Beamish D., 1986a, Deep crustal geoelectric structure beneath the Northumberland Basin, Geophys. J. R. astr. Soc. **84**, 619-640.

Beamish D., 1986b, Geoelectric structural dimensions from magnetotelluric data: Methods of estimation, old and new, Geophysics **51**, 1298-1309.

Beamish, D. and Tzanis, A. 1986, High resolution spectral characteristics of the Earth-ionosphere cavity resonances, J. atmos. terr. Phys. **48**, 187-203.

Bendat J.S. and Piersol A.G., 1971, Random Data: Analysis and Measurement Procedures, Wiley, N.Y.

Berdichevsky M.N. and Zhdanov M.S., 1984, Advanced theory of deep geomagnetic sounding, Elsevier, Amsterdam.

Bliokh H., Nikolaenko A.P., and Filippov Yu.F., 1980, Schumann resonances in the Earth-Ionosphere Cavity (LLANWYN-JONES, D., Ed.), IEE Electromagnetic Wave Series 9, P. Peregrinus Ltd., Stevenage.

Boehl J.E., Bostick Jr.F.X., and Smith H.W., 1977, An application of the Hilbert transform to the magnetotelluric method, Univ. of Texas-Austin, Engr. Research Lab.

Bott M.H., Swinburn P.M. and Long R.E., 1984, Deep stucture and origin of the Northumberland and Stainmore troughs, Proc. Yorks. geol. Soc. **44**, 479-495.

Brillinger, D.R., 1981, Time Series. Data analysis and theory, Holden-Day, San Francisco.

Burg, J.P., August 1968, A new analysis technique for time series data, paper presented at the NATO Advanced Study Institute on Signal Processing with Emphasis on Underwater Acoustics, Enschede, The Netherlands.

Burg, J.P., 1975, Maximum Entropy Spectral Analysis, Ph.D. Thesis, Stanford University, Stanford, California.

Cagniard L., 1953, Basic theory of the magnetotelluric method of geophysical prospecting, Geophysics **18**, 605-635.

Cantwell, T., 1960, Detection and analysis of low frequency electromagnetic signals, Ph.D. thesis, MIT.

Claerbout F.J., 1976, Fundamentals of Geophysical Data Processing with Applications to Petroleum Prospecting, McGraw-hill.

Clark I., 1977, ROKE, a computer program for nonlinear least-squares decomposition of mixtures of distributions, Computers and Geosciences 3, 245-256.

Clark M.W., 1977, GETHEN: A computer program for the decomposition of mixtures of two normal distributions by the method of moments, Computers and Geosciences 3, 257-267.

Clay C.S. and Hinich M.J., 1981, Estimating the Earth's impedance function when there is noise in the electric and magnetic signals, Applied Time Series Analysis II, (Findley D. F. editor), Academic Press.

Counil J.L., Le Mouël J.L. and Menvielle M., 1986, Associate and conjugate direction concepts in magnetotellurics, Annales Geophysicae 4, series B, 115-130.

Cox C.S., Filloux J.H., Gough D.I., Larsen J.C., Poehls K.A., von Herzen R.P. and Winter R., 1980, Atlantic lithospheric sounding, in Electromagnetic induction in the Earth and Moon, Schmucker, U. (editor), pp. 13-22, Centr. Acad. Publ. Japan, Tokyo and Reidel, Dordrecht.

Dubrovsky V.G. and Kramarenko S.A., 1977, Stochastic model of the Earth's electromagnetic field, Acta Geodaet. Geophys. et Montanist. Acad. Sci. Hung. **12**, 335-344.

Egbert G.D. and Booker J.R., 1986, Robust estimation of geomagnetic transfer functions, Geophys. J. R. astr. Soc. **87**, 173-194.

Eggers D., 1982, An eigenstate formulation of the magnetotelluric impedance tensor, Geophysics **47**, 1204-1214.

Fischer G. and Schnegg P.A., 1980, The dispersion relations of the magnetotelluric response and their incidence on the inversion problem, Geophys. J. R. astr. Soc. **62**, 661-673.

Fischer G., 1982, Magnetotelluric observational techniques on land, Geophysical Surveys 4, 373-393.

Fougere P.F., Zawalick E.J., and Radoski H.R., 1976, Spontaneous Line Splitting in Maximum Entropy Power Spectrum Analysis, Phys. Earth Planet. Interiors **12**, 201-207.

Fougere P.F., 1977, A Solution to the Problem of Spontaneous Line Splitting in Maximum Entropy Power Spectrum Analysis, J. Geophys. Res. **82**, 1051-1054.

Frohlich R.K., 1971, The Influence of Industrial Stray Currents on the Measurement of Earth Potentials and their Elimination, Geophysical Prospecting **19**, 118-132.

Gamble T.D., Goubau W.M and Clarke J., 1979a, Magnetotellurics with a remote magnetic reference, Geophysics **44**, 53-68

Gamble T.D., Goubau W.M and Clarke J., 1979b, Error analysis for remote reference magnetotellurics, Geophysics **44**, 959-968.

Gell'mann M. and Nee'man Y., 1965, The eightfold way, Benjamin, New York.

Gersch W. and Sharpe D.R., 1973, Estimation of power spectra with finite order autoregressive models, IEEE Trans. Automat. Contr., **AC-18**, 367-369.

Ghausi M.S. and Kelly J.J., 1968, Introduction to Distributed Parameter Networks, Holt Rinehart and Winston, New York.

Golub G., 1965, Numerical Methods for Solving Least Squares Problems, Numerische Mathematik **7**, 206-216.

Golub G. and van Loan C.F., 1983, Matrix Computations, North Oxford Academic, Oxford.

Goodman N.R., 1965, Measurement of Matrix Frequency Response Functions and Multiple Coherence Functions, AFFDL TR 65-56, Air Force Flight Dynamics Laboratory, Wright-Patterson AFB, Ohio.

Gutowski P.R., Robinson E.A. and Treitel S., 1978, Spectral Estimation: Fact or Fiction?, IEEE Trans. on Geosciences Electronics **GE-16**, no 2, 80-84.

Hampel F.R., Ronchetti E.M., Rousseeuw P.J. and Stahel W.A., 1986, Robust Statistics: The Approach Based on Influence Functions, Wiley, N.Y.

Hermance J.F., 1982, The asymptotic response of three-dimensional basin offsets to magnetotelluric fields at long periods: The effects of current channeling, Geophysics **47**, 1562-1573.

Hobbs B.A., 1982, Automatic Model for Finding the One-Dimensional Magnetotellutic Problem, Geophys. J. R. astr. Soc. **68**, 253-264.

Hohmann G.W., 1983, Three-dimensional EM Modeling, Geophysical Surveys **6**, 27-53.

Huber P.J., 1981, Robust Statistics, Wiley, New York.

Ingham M.R. and Hutton V.R.S., 1982a, Crustal and upper mantle electrical conductivity structure in Southern Scotland, Geophys. J. R. astr. Soc. **69**, 579-594.

Ingham M.R. and Hutton V.R.S., 1982b, The interpretation and tectonic implications of the geoelectric structure of Southern Scotland, Geophys. J. R. astr. Soc. **69**, 595-606.

Jackson J.D., 1975, Classical Electrodynamics. Wiley, New York.

Janoth W., 1984, Tranformation magnetotellurischer Messdaten in die Form von Seismogrammen, Protokoll uber das 10. Kolloquium 'Electro-magnetische Tiefenforschung', Grafrath/Oberbayern.

Jaynes E.T., 1963, New engineering applications of information theory, Proccedings of the First Symposium on Engineering Applications of Random Function Theory and Probability, (J.L. BOGDANOFF and F. KOZIN editors), 163-203, Wiley, New York.

Jaynes E.T., 1968, Prior Probabilities, IEEE Trans. Systems Sci. Cybern., **SEC-4**, 227-241.

Jaynes E.T., 1982, On the Rationale of Maximum Entropy Methods, Proc. IEEE **70**, no 9, 939

Jones A.G., Olafsdottir B., and Tiikkainen J., 1983, Geomagnetic induction studies in Scandinavia III. Magnetotelluric observations, J. Geophys. **54**, 35-50.

Jones F.W. and Kelly M., 1966, Man-made Telluric Micropulsations, Canadian Journal of Physics **44**, 3025-3031.

Karmann R., 1977, Search-coil magnetometers with optimum signal-to-noise ratio, Acta Geodaet. Geophys. et Montanist. Acad. Sci. Hung. **12**, 353-357.

Keller G.V. and Frischknecht F.C., 1966, Electrical Methods in Geophysical Prospecting, Pergamon Press, Oxford.

Kemp D.T., 1971, The global location of large lightning discharges from single station observations of ELF disturbances in the Earth-ionosphere cavity, J. atmos. terr. Phys. **33**, 919-927.

Kendal M.G. and Stuart A., 1967, The Advanced Theory of Statistics, Vol II, Griffin, London.

Khachay O.A., 1978, On solving the inverse problem of Magnetotelluric Sounding for a complex impedance, Phys. Solid Ear. Izvestiya **14**, 896-900.

Kröger P, Micheel H.J. and Elsner R., 1983, Comparison of Errors in Local and Reference Estimates of the Magnetotelluric Impedance Tensor, J. Geophys. **52**, 97-105.

Kromer R., 1970, Asymptotic properties of the autoregressive spectral estimator, Ph.D. thesis, Dep. of Statistics, Stanford Univ., Stanford, Calif.

Kunetz G., 1972, Processing and Interpretation of Magnetotelluric Soundings, Geophysics **37**, 1005-1021.

Lanzcos C., 1961, Linear differential operators. Van Nostrand Ltd, Princeton.

Larsen J.C., 1975, Low Frequency (0.1-6.0cpd) Electromagnetic Study of Deep Mantle Electrical Conductivity Beneath the Hawaiian Islands, Geophys. J. R. astr. Soc. **43**, 17-46.

LaTorraca G.A., Madden T.R. and Korringa J., 1986, An analysis of the magnetotelluric impedance for three-dimensional conductivity structures, Geophysics **51**, 1819-1829.

Lawson C.L. and Hanson R.J., 1974, Solving least squares problems, Prentice-Hall, Englewood Cliffs.

Lee S., McMechan G.A. and Aiken C.L.V, 1987, Phase field imaging: The electromagnetic equivalent of seismic migration, Geophysics **52**, 678-693.

Legget J.C., McKerrow W.S. and Soper N.J., 1983, A model for the crustal evolution of southern Scotland, Tectonics **2**, 187-210.

Lokken J.E. and Shand J.A., 1964, Man-made electromagnetic interference at extremely low frequencies, Canadian Journal of Physics **42**, 1902-1907.

Malin S.R.C., Barraclough D.R. and Hodder B.M., 1982, A Compact Algorithm for the Formation and Solution of Normal Equations, Computers and Geosciences **8**, 355-358.

Menvielle M. and Szarka L., 1986, Distortions of electromagnetic fields: Topographic and man-made. Review paper presented in the VIII IAGA Workshop on Electromagnetic Induction, Neuchatel, Switzerland, August 1986.

Morf M., A. Vieira D.T.L. Lee and T. Kailath, 1978, Recursive multichannel Maximum Entropy spectral estimation, IEEE Trans. on Geoscience Electronics **GE-16**, no 2, 85-94.

Murnaghan F.D., 1938, The Theory of Group Representations, Johns Hopkins Press, Baltimore.

Newton R.G., 1984, An inverse spectral problem in three dimensions, SIAM-AMS Proc. **14**, 81-90.

Normand J.M., 1980, A Lie Group: Rotations in Quantum Mechanics, North-Holland, Amsterdam.

Park S.K., Orange A.S. and Madden T.R., 1983, Effects of three-dimensional structure on magnetotelluric sounding curves, Geophysics **48**, 1402-1405.

Park J. and Chave A.D, 1984, On the estimation of Magnetotelluric Response Functions using the Singular Value Decomposition, Geophys. J. R. astr. Soc. **77**, 683-709.

Parker R.L., 1980, The Inverse Problem of Electromagnetic Induction: Existence and Construction of Solutions Based On Incomplete Data, J. Geophys. Res. **85**, 4421-4428.

Parker R.L. and Whaler K.A., 1981, Numerical methods for establishing solutions to the inverse problem of Electromagnetic Induction, J. Geophys. Res. **86**, 9574-9584.

Parker R.L., 1983, The Magnetotelluric Inverse Problem, Geophysical Surveys **6**, 5-25.

Peacock K.L., and Treitel S., 1969, Predictive deconvolution: theory and practice, Geophysics **34**, 155-169.

Pedersen L.B., 1982, The magnetotelluric impedance tensor - its Random and Bias errors, Geophysical Prospecting **30**, 188-210.

Pedersen L.B., and Svennekjaer M., 1984, Extremal bias coupling in magnetotellurics, Geophysics **49**, 1968-1978.

Porstendorfer G., 1975, Principles of Magneto-Telluric Prospecting, Gepexploration Monographs, Ser. 1, No 5, Gebruder-Borntraeger, Berlin W.

Press W.H., Flannery B.P., Teukolsky S.A. and Vetterling W.T., 1986, Numerical Recipes: The Art of Scientific Computing, Cambridge University Press, Cambridge.

Priestley M.B., 1981, Spectral Analysis and Time Series - Volume 2: Multivariate series, Prediction and Control, Academic Press, London.

Raiche A.P., 1973, An integral equation approach to three-dimensional modelling, Geophys. J. R. astr. Soc. **36**, 363-376.

Reddy I.K. and Rankin D., 1974, Coherence functions for magnetotelluric analysis, Geophysics **39**, 312-320.

Robinson E.A., 1962, Random Wavelets and Cybernetic Systems, Griffin, London.

Robinson E.A., 1967, Multichannel Time Series Analysis with Digital Computer Programs, Holden-Day, San Francisco.

Robinson E.A., 1980, Physical Applications of Stationary Time Series, Griffin, London.

Rokityansky I.I., 1982, Geoelectromagnetic Investigation of the Earth's Crust and Mantle, Springer-Verlag.

Rose M.E., 1957, Elementary theory of angular momentum, Wiley, New York.

Schnegg P.A., Fischer G., Fontes S., Hutton V.R.S. and Finzi E., The Effect of D.C. Railways on MT Measurements in Northern Italy, paper presented in the VIII IAGA Workshop on Electromagnetic Induction, Neuchatel, Switzerland, August 1986.

Sims W.S., Bostick Jr.F.X. and Smith H.W., 1971, The estimation of magnetotelluric impedance tensor elements from measured data, Geophysics **36**, 938-942.

Spitz S., 1985, The magnetotelluric impedance tensor properties with respect to rotations, Geophysics **50**, 1610-1617.

Stodt J.A., Weighted Averaging and Coherence Sorting for Least Squares Magnetotelluric estimates, paper presented in the VIII IAGA Workshop on Electromagnetic Induction, Neuchatel, Switzerland, August 1986.

Strand O.N., 1977, Multichannel complex maximum entropy (Autoregressive) spectral analysis, IEEE Trans. on Autom. Control, **AC-22**, no 4, 634-640.

Svetov B.S., Kurginyan S.E. and Shimelevitch M.I., 1978, About interpretation of frequency characteristics of electromagnetic fields in the range of complex frequency (in Russian), Izv. Akad. Nauk SSSR Fiz. Zemli **6**, 50-56.

Swift C.M. Jr, 1967, A magnetotelluric investigation of an electrical conductivity anomaly in the southwestern United States, Ph.D. thesis, MIT.

Ting S.C. and Hohmann G.W., 1981, Integral equation modeling of three-dimensional magnetotelluric response, Geophysics **46**, 182-197.

Tzanis A. and Beamish D. 1987a, Time domain polarization analysis of Schumann resonance waveforms, J. Atmos. terr. Phys. **48**, 217-229.

Tzanis A. and Beamish D., 1987b, Audiomagnetotelluric Crustal Sounding Using the Schumann Resonanses, J. Geophys. **61**, 97-100.

Ulrych T.J. and Bishop T.N., 1975, Maximum Entropy Spectral Analysis and Autoregressive Decomposition, Rev. geophys. space Phys. **13**, 183-200.

Ulruch T.J. and Jensen O.G., 1974, Cross-spectral analysis by using Maximum Entropy, Geophysics **39**, 353-354.

Ward S.H., Peeples W.J. and Ryu J., 1973, Analysis of Geoelectromagnetic Data, in Methods in Computational Physics (B.A. Bolt edt) v. 13, Academic Press, New York.

Weidelt P., 1972, The Inverse Problem of Geomagnetic Induction, Zeitschrift für Geophysik **38**, 257-289.

Weidelt P., 1975, Electromagnetic Induction in Three-Dimensional Structures, J. Geophys. **41**, 85-109.

Weidelt P., 1985, Construction of conductance bounds from magnetotelluric impedances, J. Geophys. **57**, 191-206.

Widrow B., McCool J.M., Larimore M.G. and Johnson C.R., Jr., 1976, Stationary and Non-stationary characteristics of the LMS adaptive filter, Proc. IEEE **64**, 1151-1162.

Wight, D.E., Bostick F.X.Jr, and Smith W.H., 1977, Real time Fourier transformation of magnetotelluric data, Electrical Geophysics Laboratory, University of Texas at Austin.

Wigner E.P., 1959, Group theory and its applications to the quantum mechanics of atomic spectra, Academic Press, New York.

Word D.R., Smith H.W. and Bostick F.X. Jr., 1970, An investigation of the magnetotelluric tensor impedance method. Tech. rep. 82, Electr. Geophys. Res. Lab., Univ. of Texas at Austin.

Zhang P., Roberts R.G. and Pedersen L.B., 1987, Magnetotelluric strike rules, Geophysics **52**, 267-277.

APPENDIX I

A STUDY OF THE EARTH-IONOSPHERE CAVITY (SCHUMANN) RESONANCES

-AN IMPORTANT AMT SOURCE FIELD-

Journal of Atmospheric and Terrestrial Physics, Vol. 48, No. 2, pp. 187–203, 1986.
Printed in Great Britain.

0021-9169/86 $3.00 + .00

High resolution spectral characteristics of the Earth–ionosphere cavity resonances

D. BEAMISH and A. TZANIS*

British Geological Survey, Murchison House, West Mains Road, Edinburgh EH9 3LA, U.K.

(*Received in final form* 24 *June* 1985)

Abstract—The natural resonances of the Earth–ionosphere cavity at frequencies between 5 and 100 Hz have been studied since the fundamental paper by Schumann. While the gross features of the phenomena are now well understood, considerable work remains to be done on their detailed behaviour. In the present study a high resolution, data adaptive spectral technique is applied to digital electromagnetic data obtained at a moderate latitude. A particular feature of the method employed is that spectral properties become available on the same time scale as many ELF events, thus both time local and time averaged resonance features can be readily established. The technique can thus be applied to both dynamic and steady-state descriptions of the cavity's properties.

For the data set considered, the technique adequately resolves the first six resonance modes on a time scale of 0.75 s. The presence of higher order modes is also indicated. The time averaged frequencies obtained are in accord with those of previous experimental determinations. When the time local properties of individual transient waveforms are examined, however, we observe a number of detailed effects which are predicted by theory. The precise spectral structure of the resonance modes appears influenced by the differing locations of the sources of transient excitation. In the case of the first order resonance mode, the properties of the cavity consistently support both singlet and doublet resonance behaviour.

1. INTRODUCTION

ELF waves in the Earth–ionosphere waveguide, generated by the global sum of thunderstorm activity, are able to propagate over large distances with very little attenuation (GALEJS, 1964). The waveguide becomes a cavity resonator at such frequencies and for a perfectly reflecting ionosphere the first five resonance modes would be established at frequencies of 10.6, 18.4, 26.0, 35.5 and 41.1 Hz (MADDEN and THOMSON, 1965). In reality, due to energy losses, the first five modes have typical lower frequencies observed at 7.8, 14.1, 20.3, 26.3 and 32.5 Hz (MADDEN and THOMSON, 1965). These are termed the Schumann Resonances. As pointed out by GALEJS (1961) and by POLK (1969), the relative amplitudes of the resonance peaks in the Schumann power spectrum depend primarily on the source–receiver separation (θ), while the precise values of the resonance frequencies depend not only on θ, but also on the electrical conductivity of the ionosphere. Thus it is anticipated that the analysis of the Schumann spectrum will permit a description of the world-wide average conductivity profile of the lower ionosphere (TRAN and POLK, 1979a).

The resonant eigenfrequencies established from theory are a function of the complexity of the model considered. In models which treat the ionosphere as spherically symmetric, the eigenvalues of the wave equation are degenerate, giving rise to single, precise eigenfrequencies for the resonant modes. Such degeneracy may be partially or completely removed when azimuthal asymmetries in the waveguide (e.g. the day–night ionospheric inhomogeneity) are taken into account. Under such circumstances, frequency splitting of the fundamental resonant modes may occur (MADDEN and THOMSON, 1965). An additional complication is introduced when the Earth's magnetic field is taken into account. The presence of the magnetic field necessitates the inclusion of gyrotropic waveguide properties and results in different phase velocities for waves travelling in opposite directions around the Earth (BLIOKH *et al.*, 1980). This again has the resultant effect of producing frequency splitting of the resonant modes. In the mathematical development of the field expressions, the Legendre polynomials of degree n are replaced by the associated Legendre polynomials of degree $(2n+1)$, causing a $(2n+1)$ splitting of the resonant frequency.

BLIOKH *et al.* (1980) demonstrate such effects of the first ($n = 1$) resonant mode by considering three models. Model (a) consists of a perfect Earth resonator, model (b) consists of an Earth resonator with polar inhomogeneity and model (c) consists of an Earth resonator with a dipole magnetic field. For model (a)

*Also at: Geophysics Department, University of Edinburgh, King's Buildings, J.C.M.B. Mayfield Road, Edinburgh, U.K.

the first mode has the degenerate frequency of $f = 10.6$ Hz. In model (b) the waveguide can support oscillations at two frequencies $f(1) = 8$ Hz ($m = 0$) and $f(2) = 7.8$ Hz ($m = 1$), where m is the azimuthal wave velocity of the wave assumed to possess an exp (imQ) azimuthal dependence. In model (c) the $(2n+1)$ triplet split is observed with the frequencies established at $f(1) = 7.5$ Hz ($m = -1$, west to east equatorial propagation), $f(2) = 8$ Hz ($m = 0$) and $f(3) = 8.16$ Hz ($m = +1$). The model parameters were chosen to produce a normalized frequency of 8 Hz for the $m = 0$ mode.

The above models constitute 'steady-state' descriptions of the waveguide properties in which the ionospheric conductivity profile is essentially fixed. Since the ionosphere has dynamic properties, a number of papers detail how the electrical conductivity profile of the lower ionosphere (*C*-, *D*- and *E*-regions) affects the characteristics of the Schumann resonance spectra (MADDEN and THOMSON, 1965; TRAN and POLK, 1979b; SENTMAN, 1983). TRAN and POLK (1976) have shown that the region which determines the resonance frequencies is between 40 and 100 km. Different ionisation mechanisms are responsible for maintaining the electrical conductivity at different altitudes through this interval. Ionisation sources include galactic cosmic rays (lower *D*- or *C*-region), solar Lyman-α (*D*-region) and solar X-rays (*E*-region). Changes in the intensity or availability of such sources therefore modulate the ionospheric conductivity at the corresponding altitude. These variations in turn influence the characteristics of the Schumann resonances. Typically, an increase of conductivity below 63 km (lower *D*-region acting as a dielectric) would lower the resonance frequencies, while an increase of conductivity within the upper *D*-region (acting as an imperfect conductor) would lead to an increase in the resonance frequencies. In the model ionosphere studied by TRAN and POLK (1979a), changes in the conductivity of the *D*- and lower *E*-regions led to frequency shifts of 10% or more in the first three resonance frequencies. TRAN and POLK (1979b) provide a method for evaluating conductivity profiles based on three sets of observational parameters: (a) the exact values of the resonance frequencies; (b) the ratios of the spectral amplitudes at successive resonance peaks; (c) the bandwidths of the resonance peaks at the 90% amplitude level.

SENTMAN (1983) has studied the effects on the Schumann resonance frequencies of perturbations in a 'steady-state' exponential ionospheric profile as an aid in interpreting observations in terms of the sources of ionisation. The models used provide overall frequency variations of ± 1.0, ± 1.5 and ± 2.5 Hz for the $n = 1, 2$ and 3 Schumann resonance modes. In addition, fluctuations in the resonance frequencies from observational data have been studied in relation to the occurrence of sudden ionospheric disturbances, having onset times of the order of minutes (CANNON and RYCROFT, 1982). In all cases it is necessary to establish precise values of the resonance frequencies from experimental data sets over appropriate time windows.

The ELF waves propagate as a 0th order TM mode in which the electric field is largely radial. As a consequence, vertical antennas are often used as appropriate omnidirectional receivers of the ELF noise in the audio and subaudio bands. Equally, sensitive magnetometers, such as induction coils, measuring the horizontal components of the magnetic field may also be used to provide equivalent data. Such data sets have been used to estimate resonance frequencies with a temporal resolution (integration time) of 13.6 min (CANNON and RYCROFT, 1982) and 17 and 34 min (TRAN and POLK, 1979b). Such a long time base appears to be required due to the Fourier techniques used to analyse the data sets. Using such data and integration times, the typical frequency resolution is quoted as being 0.156 Hz (CANNON and RYCROFT, 1982) and 0.125 Hz (TRAN and POLK, 1979b).

Clearly, any improvement in time and/or frequency resolution would assist in the comparison of models and observations. The purpose of the present paper is to point out the growing availability of digital audio frequency data sets and to demonstrate a high time/frequency resolution spectral technique that can be applied in the analysis of such data. The technique permits adequate frequency resolution over time windows of several seconds for the $n = 1$ mode and less than one second for the higher modes. Since such time windows are of the same order as the duration of many transient ELF 'events' (JONES and KEMP, 1971; BLIOKH *et al.*, 1980), the dynamic components of the resonance phenomena are highly resolved and the average, or integrated, properties can be established across relatively short time periods.

2. AUDIOMAGNETOTELLURIC DATA

Audiomagnetotelluric (AMT) geophysical investigations provide a means of studying the Earth's electrical conductivity structure at shallow crustal depths. Methodologies are described by KELLER (1971), STRANGWAY *et al.* (1973) and FISCHER (1982). The basic data set collected at a given location consists of two (usually orthogonal) horizontal telluric channels, measuring time changes in the induced electric field, and two components of the horizontal magnetic field. The bandwidth available is governed by the magnetic sensors employed in the field instrumentation. AMT

Table 1. Frequency characteristics of the three decade data scheme used in the present study

	f_L	f_H	Δf
Decade 1	10	100	400
Decade 2	1	10	40
Decade 3	0.1	1	4

f_L, low frequency −3 dB point; f_H, high frequency −3 dB point; Δf, data sampling frequency. All frequencies in Hz.

instrumentation nominally covers the frequency range 1–100 Hz. However, the high frequency limit may extend from 10 to 100 KHz.

The data described in the present study were acquired with a minicomputer based AMT field system covering the bandwidth 0.01–100 Hz. Band-pass and notch filters are applied to condition the analogue signals prior to 12-bit digitisation. The conditioning scheme provides the 3 decades and sampling rates shown in Table 1. In relation to the Schumann resonances, decade 2 provides data for studies of the $n = 1$ (8 Hz) mode, while decade 1 provides data for studies of the higher order modes. The two decades are not recorded simultaneously. The four or more data channels are collected across a window of 300 data points. The time interval for the collection of each data window is thus 0.75 s for decade 1 and 7.5 s for decade 2. Data collection between successive data windows is at present discontinuous due to in-field processing and data storage on a slow digital cartridge. Such discontinuities are not a limitation to the subsequent spectral analysis, since the 300 data points sampled at twice the Nyquist frequency provide sufficient degrees of freedom across the bandwidth of each decade. Such temporal resolution is advantageous only if spectral characteristics can be adequately resolved from the number of data points made available in each data window. This question is considered in the following section.

When examining the data on these time scales it is possible that either the resonance waveforms are totally absent, for certain data windows, or that the waveforms exhibit large amplitude variance over successive data windows. Such amplitude variance is common. An example is provided in Fig. 1, which displays decade 1 data (10–100 Hz) for five successive data windows in two orthogonal telluric (induced electric) and two orthogonal magnetic channels.

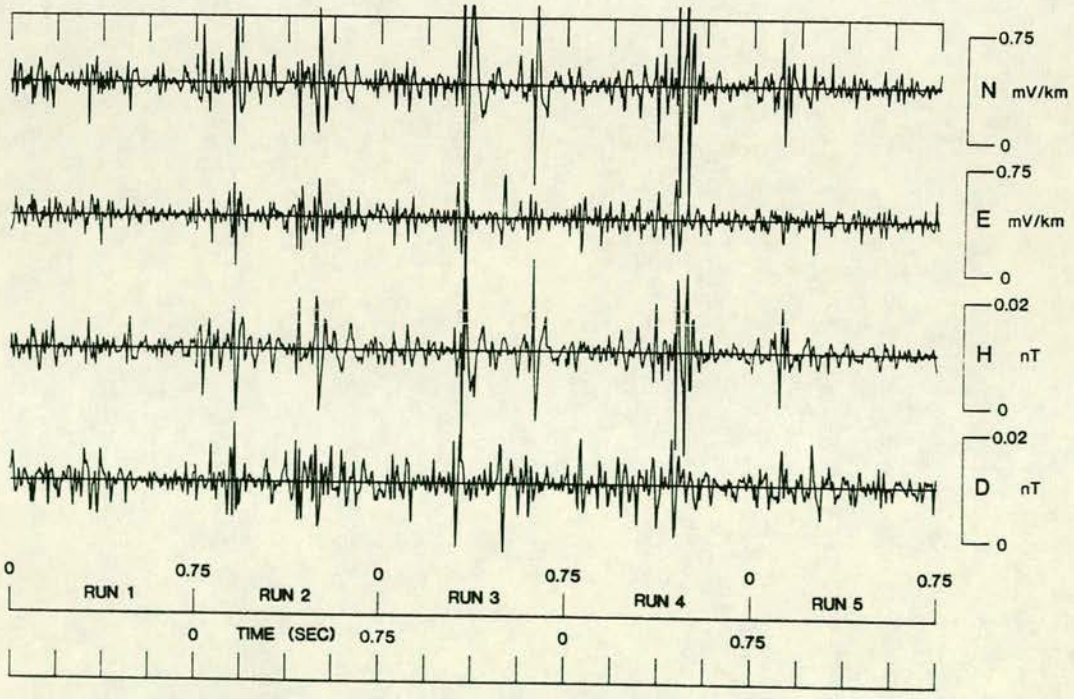


Fig. 1. Four data channels of decade 1 data. Sampling interval on plot is 200 Hz, obtained from data collected at 400 Hz. Five successive data windows of 0.75 s length. *N* and *E* refer to 2 orthogonal components of the induced electric field. *H* and *D* are the corresponding magnetic N–S and E–W components.

For a single radial source filament, the horizontal magnetic field ($H_{\phi n}$) generated is proportional to the angular source–receiver separation (GALEJS, 1972) as

$$H_{\phi n} \propto \frac{\partial}{\partial \theta} P_n(\cos\theta),$$

where the subscript n is the resonance mode number and $P_n(\cos\theta)$ is the Legendre function of order n. In spectral analysis of any given data window only power magnitudes are required, since the cavity resonance may be excited simultaneously by a multiplicity of incoherent sources. For each thunderstorm centre considered, the lightning spectrum may be represented by the relation (GALEJS, 1961)

$$g(\omega) = c \times \exp(-9.1 \times 10^{-3}\,\omega)$$

with c a constant and ω angular frequency. The received power spectrum, $G_n(\omega)$, will be (POLK, 1969)

$$G_n(\omega) = \int_{\phi_1}^{\phi_2} \int_{\theta_1}^{\theta_2} g(\omega)|H_{\phi n}|^2 \sin\theta \, d\theta \, d\phi$$

for a uniform distribution of $g(\omega)$ between the limits θ_1, θ_2 and ϕ_1, ϕ_2. For multiple and distributed thunderstorm centres, the power spectrum received is thus the sum of individual contributions of the form given above. The length of the data windows used in the present study are of the same order as the time scales of the models of lightning discharge (e.g. JONES and KEMP, 1971) and we therefore anticipate large variations in signal/noise ratios.

The data examined in the present study were collected in western Anatolia (Turkey), geographic coordinates 40.5°N, 30°E. The data were selected purely on the grounds of good signal/noise ratios. The decade 1 and decade 2 data were collected on days 209 and 159, 1984, respectively.

3. DATA ANALYSIS

Figure 2 shows typical data recorded in orthogonal electric (E) and magnetic (H) channels and the associated raw (unsmoothed) power spectra for decade

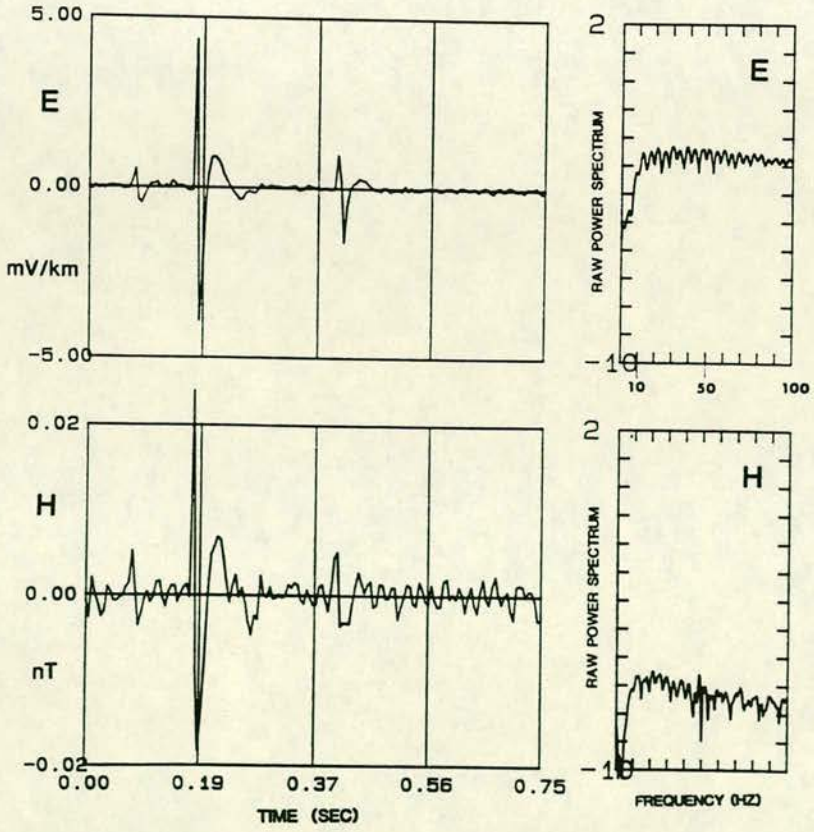


Fig. 2. Data and raw power spectra from two channels of decade 1 data. E is the E–W induced electric field, H is the magnetic N–S component. Raw spectra in units of $(\text{mV km}^{-1})^2\,\text{Hz}^{-1}$ (E) and $(\text{nT})^2\,\text{Hz}^{-1}$ (H).

1 data. The relatively flat power spectra observed indicate a white noise process. In calculating the power spectrum, and associated confidence limits, using the discrete Fourier transform, it is necessary to smooth adjacent power estimates to reduce variance. If it is assumed that the data consist of a white noise series ($n(i)$) of length $N \cdot \Delta t$, the fractional error ε (the ratio of RMS deviation to mean) in the smoothed spectral estimates is given by

$$\varepsilon = \frac{1}{\sqrt{2m+1}}$$

when the raw estimates are averaged over $2m+1$ data points (JENKINS and WATTS, 1968). For a given fractional error, the raw spectral estimates must be smoothed over the frequency interval Δf given by

$$\Delta f = \frac{1}{2 \cdot N \cdot \Delta t \cdot \varepsilon^2}.$$

Thus for a stable estimate of the power spectrum, ε must clearly be small and the estimate must be smoothed over a wide frequency range. The reliability of the estimate is improved at the expense of frequency resolution unless N is made large.

If the time series is of autoregressive form, i.e.

$$n_i + \sum_{m=1}^{M} a_m n_{i-m} = x_i,$$

where x_i is a white noise series, it is possible to improve the resolution beyond the above limits for a given N. The property of the filter ($a(m), m = 1, M$) that generates the autoregressive time series is determined by a small number of filter coefficients. The filter coefficients can be determined using a finite portion of the autocorrelation function $R_m = \langle n(i)n(i+m) \rangle$ for $m = 0, M-1$ and where the brackets denote ensemble average. The power spectrum can then be determined from the filter coefficients obtained. This is the basis of the maximum entropy method of spectral analysis (e.g. ULRYCH and BISHOP, 1975).

A number of methods exist for estimating the autoregressive (AR) filter coefficients. Among these the algorithm due to BURG (1968) is both data adaptive and computationally efficient. The highest order filter coefficient (reflection coefficient) is determined by minimizing the error power output from the filter fitted to the data in both time directions. The remaining coefficients are calculated using the LEVINSON (1947) recursive algorithm, which involves previously determined lower order filters. The procedure is iterative and fits successively higher order autoregressive operators to the data until a specified order is reached or a cut-off criterion is satisfied. This procedure ensures that the filter is minimum delay, and therefore stable. When referring to ME spectra we henceforth implicitly refer to Burg's algorithm.

Once the AR coefficients are determined, the power spectrum can be obtained through the inverse of the z-transform of the filter scaled by the minimum residual power of the input time series. The advantage of the power spectrum so derived is the inherent frequency resolution, which is superior to any of the standard spectral analysis techniques. It is worth noting that in the case of ME spectra, it is the area under a given spectral peak that determines the power level, rather than the actual peak height.

A major problem that arises in the use of the ME method is the determination of the order of the AR process, i.e. the number of filter coefficients (M) that are sufficient to describe the process. If M is too small, the data are underfitted, a smooth spectrum will result and the high resolution capability of the method is lost. If M is too large and the data are overfitted, undesired effects may occur. This problem has led to a number of criteria to establish the correct order of the AR process. The most widely used theoretical criteria are Akaike's final prediction error (FPE) criterion, (AKAIKE, 1969, 1970), Akaike's theoretical information criterion (AIC) (AKAIKE, 1974, 1976) and Parzen's criterion for autoregressive transfer functions (CAT) (PARZEN, 1976). Other empirical criteria have been proposed where the above fail or seem to be inconclusive. The empirical criteria suggest limiting the order of the filter length (M) to a given fraction of the data length (N). ULRYCH and CLAYTON (1976) propose a filter length of between $N/2$ and $N/3$, while BERRYMAN (1978) proposes a length of $2N/\ln(2N)$.

For the present study, use was made of Akaike's FPE and AIC criteria. The behaviour of the AR filter as monitored by the performance of these criteria shows a high degree of variability between data windows. Figure 3a shows the results from six typical data windows displaying the variation of FPE and AIC, normalized to unity, against the order of the AR process, i.e. the number of filter coefficients M, where M ranges from 1 to 100. The data length in each case is $N = 150$. It can be seen that the normalized FPE and AIC parameters display a series of local minima, rather than the expected monotonically decreasing character, while the absolute minimum displays large variability in terms of the order of the AR model between the six data windows.

Because of the poor performance of the theoretical criteria, the selection of the order of the AR model was necessarily carried out semi-empirically. Clearly, the order of filter that will adequately resolve at least all the spectral peaks predicted by theory and that will not

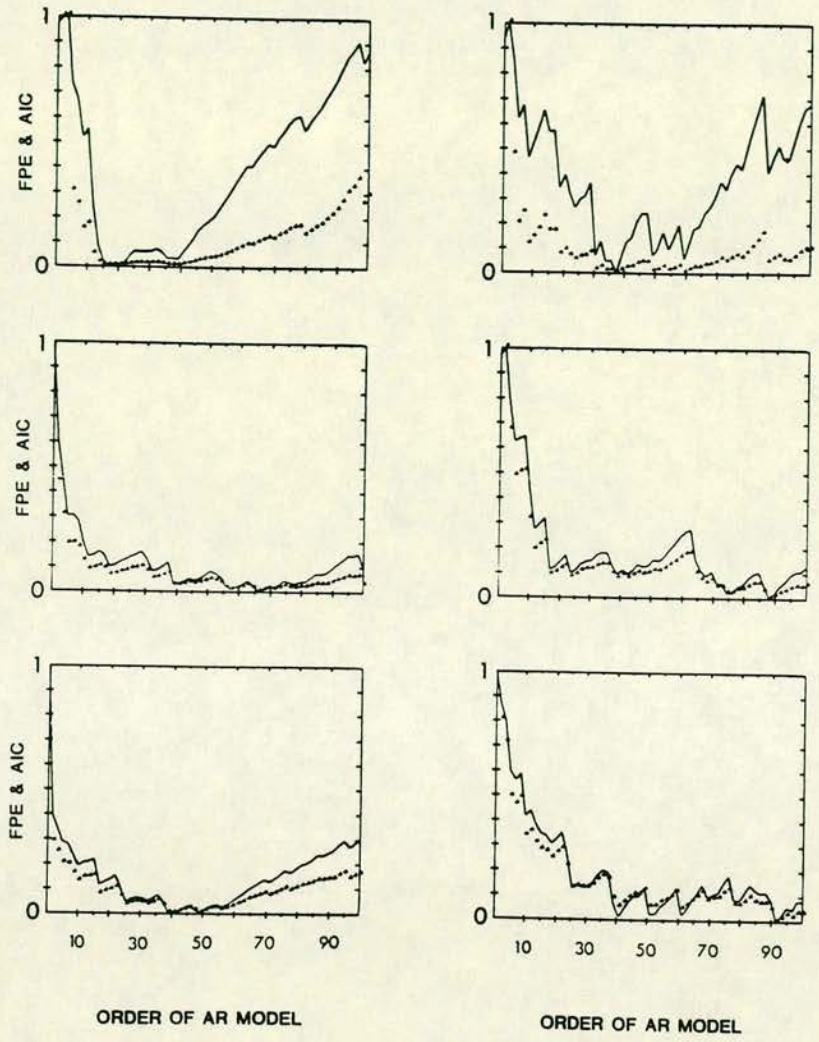


Fig. 3a. Typical performance of two criteria for estimating filter length (order of AR model) on 6 windows of H data. FPE, solid line, is Akaike's final prediction error. AIC, dotted line, is Akaike's theoretical information criterion. Values are normalized to a maximum value of unity.

overfit the data with the associated undesired effects on the majority of the data windows can be defined to be optimum for our present purposes. Given the large number of data runs available, it is possible to examine the probability distribution of the position of the absolute minimum of the FPE as a function of M, the order of the AR model. Figure 3b displays the frequency of occurrence of the absolute minimum as a function of M, using 150 data windows of decade 1 and decade 2 data. The distributions obtained display different characteristics for the two decades. For decade 2 data the position of the absolute minimum occurs for the majority of data windows in the interval $M = 10$–25. For decade 1 data the equivalent interval is $M = 25$–35. Filter lengths of $M = 12$, 18, 25 and 35 have been implemented with all data lengths set at $N = 150$. The maximum filter length used is in fact less than any of the empirical criteria quoted above. The spectral resolution available using such filter lengths is illustrated in Fig. 4. The average spectrum normalized to unity, has been obtained from 25 successive data windows using $M = 12$, 18 and 25 for decade 2 (Fig. 4a) and decade 1 (Fig. 4b). We retain a spectral resolution of 1.0 Hz (decade 1) and 0.1 Hz (decade 2), throughout. It can be seen in Fig. 4a that the 7–8 Hz peak obtained for $M = 12$ decomposes to two stable peaks for higher order lengths. In Fig. 4b it is observed that a full complement of the higher order Schumann resonance

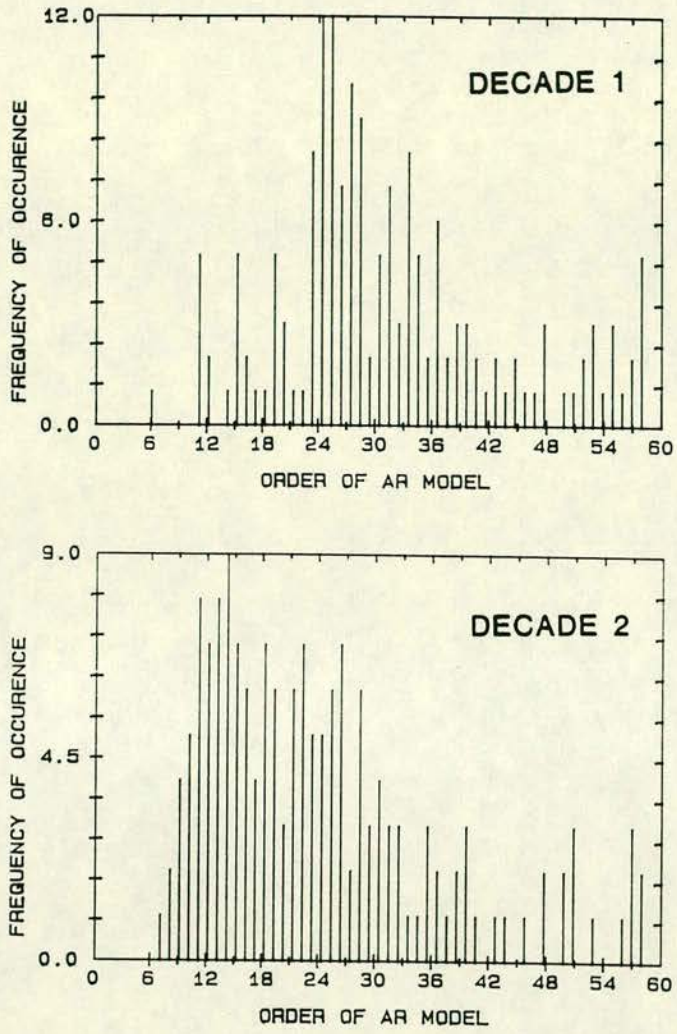


Fig. 3b. Frequency of occurrence of the position of the absolute minimum of the FPE criterion as a function of the filter length (order of AR model) using 150 data runs of decade 1 and decade 2 data. Filter lengths from 1 to 60 have been used.

modes is only established using a filter of length $M = 25$. In view of these results and those presented in Fig. 3b, filter lengths of $M > 40$ (i.e. $> 26\%$ of the data length) do not appear warranted for the present data. Specifically, for the examination of time averaged properties we consider optimum filter lengths to be $M = 18$ for decade 2 data and $M = 25$ for decade 1 data. For the examination of time local properties (e.g. sonograms) higher order filter lengths of $M = 25$ (decade 2) and $M = 35$ (decade 1) are required and have been implemented.

The term 'undesired effects' in the ME literature is understood to mean non-physical line splitting and frequency shifting of spectral peaks. Such phenomena have been the subject of rigorous investigation by several authors, as they limit the performance of the ME method for certain types of data. JAYNES (1982) has shown that such effects occur during the analysis of 'circular' time series, in which the independence of consecutive data windows is reduced by the circularity of the process. This is simulated as phase jumps that occur at regular time intervals, and is manifested by the appearance of multiple spectral lines close to the nominal frequency. Furthermore, as FOUGERE (1977) points out, such effects occur in the low noise case and are unique to the ME method. It is therefore anticipated that such problems arise during the analysis of slowly varying phenomena of low order processes with

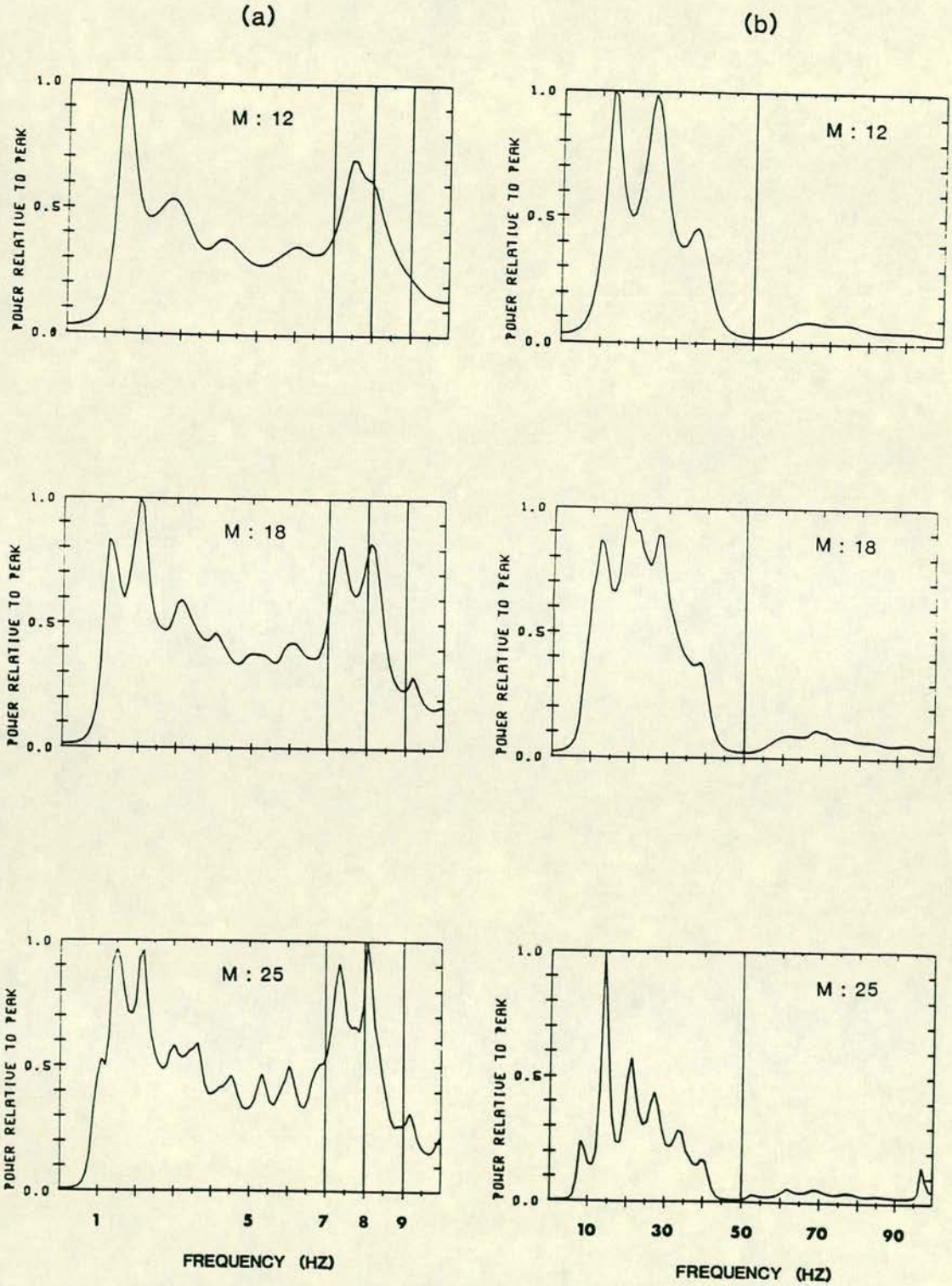


Fig. 4. Effect of a change of filter length (M) on spectral resolution. Three cases are $M = 12$, 18 and 25. (a) Decade 2 data, 1–10 Hz; (b) decade 1 data, 10–100 Hz.

moderate to large signal-to-noise ratios, such as, for instance, artificial time series of sinusoids in noise, with variable initial phases. Our data, in general, possess none of the above qualities. However, rigorous comparisons have been carried out of the peak frequencies resolved by the ME and FFT methods, with the assistance of theory for control, and we are satisfied that the ME method used in the present analysis produces no substantial undesired effects for the filter lengths implemented.

4. RESULTS

Each ME spectrum is obtained from individual data windows of 150 points. The time intervals are 0.75 s for decade 1 (10–100 Hz) and 7.5 s for decade 2 (1–10 Hz) data. Successive data windows are separated by approximately 3 s. As shown in Fig. 5, the H (N–S) magnetic component displays the largest resonance power level of the 4 component data and this component is therefore used to display the Schumann resonance characteristics.

The results are first considered as a series of evolutionary spectra from successive data windows. Each spectrum is plotted on a linear frequency scale and has been normalized to a peak value of unity to allow for the variance in waveform amplitudes. In addition, as successive data windows are accumulated, the integrated average spectrum is displayed after each accumulation of 25 spectra. Since we have not, at this point, introduced error statistics in the computation of the individual ME spectra, the accumulated average spectrum is used as a more accurate guide to the stable resonance peaks. We retain a spectral resolution of 0.1 Hz for decade 2 and 1.0 Hz for decade 1, throughout. The aim of this method is to simultaneously display time local and time averaged determinations of the resonance frequencies determined by the technique.

Figure 6 shows this form of display for decade 1 data for six successive accumulations of 25 data windows. For each set of 25 data windows (Fig. 6a–f), the 25 individual spectra are overlaid on the left and the accumulated average spectrum is shown on the right. Thus the accumulated average spectrum of Fig. 6f has

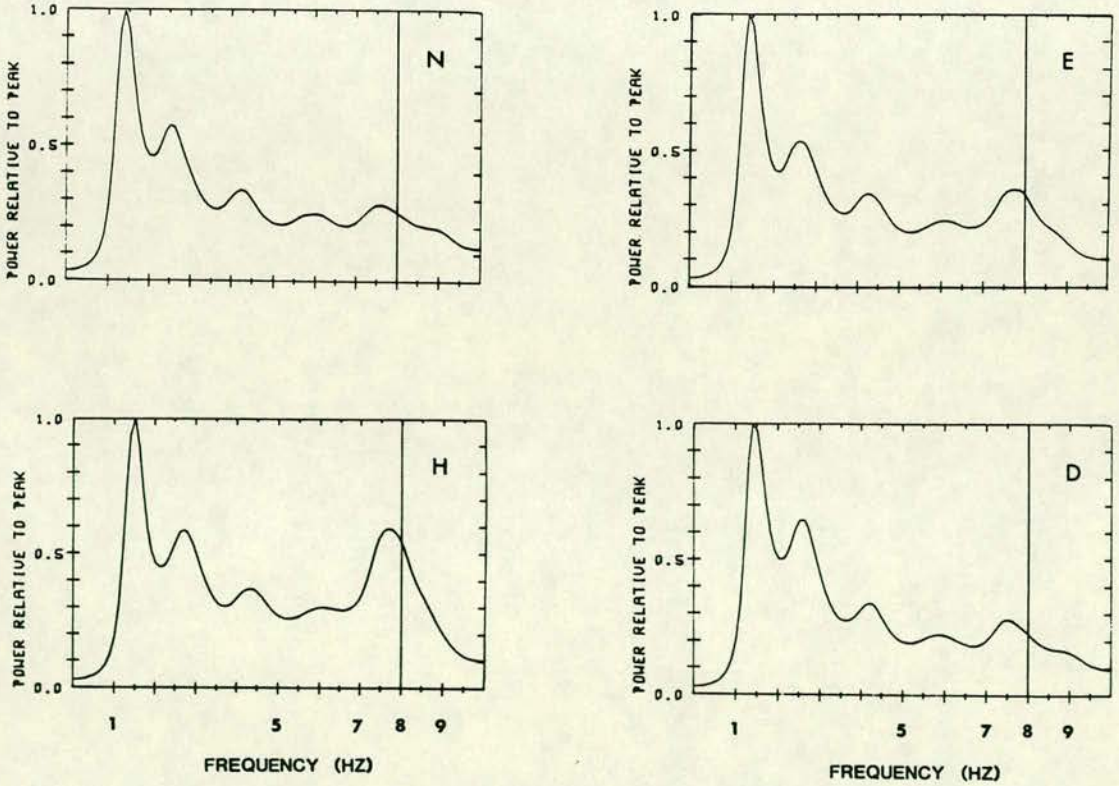


Fig. 5. Equivalent spectral peaks in the 4 component data. Average spectra using a filter length of 12, of decade 2 data. N and E, induced electric field. H and D, magnetic field components.

been obtained using 150 successive data windows. Each of the six sets of 25 data windows shown in Fig. 6 correspond to a real time interval of 94 s. A filter length $M = 25$ was used throughout. The spectra shown in Fig. 6 define the $n = 2$–6 order Schumann resonance peaks. The first order peak below 10 Hz, although present, is not quantitatively resolved due to the high pass filter stage applied to data from this decade. Stable peaks are resolved at 14.0, 21.0, 27.0, 33.5 and 39.5 Hz. The peak frequency and normalized power levels for the six accumulated average spectra are given in Table 2. The $n = 2$–5 stable peak frequencies compare well with other typical frequencies, such as 14.1, 20.3, 26.3 and 32.5 Hz (MADDEN and THOMSON, 1965).

When considering spectral details from decade 1 data some care should be exercised at frequencies of 50 Hz and above. Electromagnetic noise at such frequencies arises from the power distribution grid. The power-line voltage waveform is usually stable at 50 Hz, however, the current waveform is often complicated and highly variable due to variable loads. Generally, power-line noise consists of steady spectral lines at the

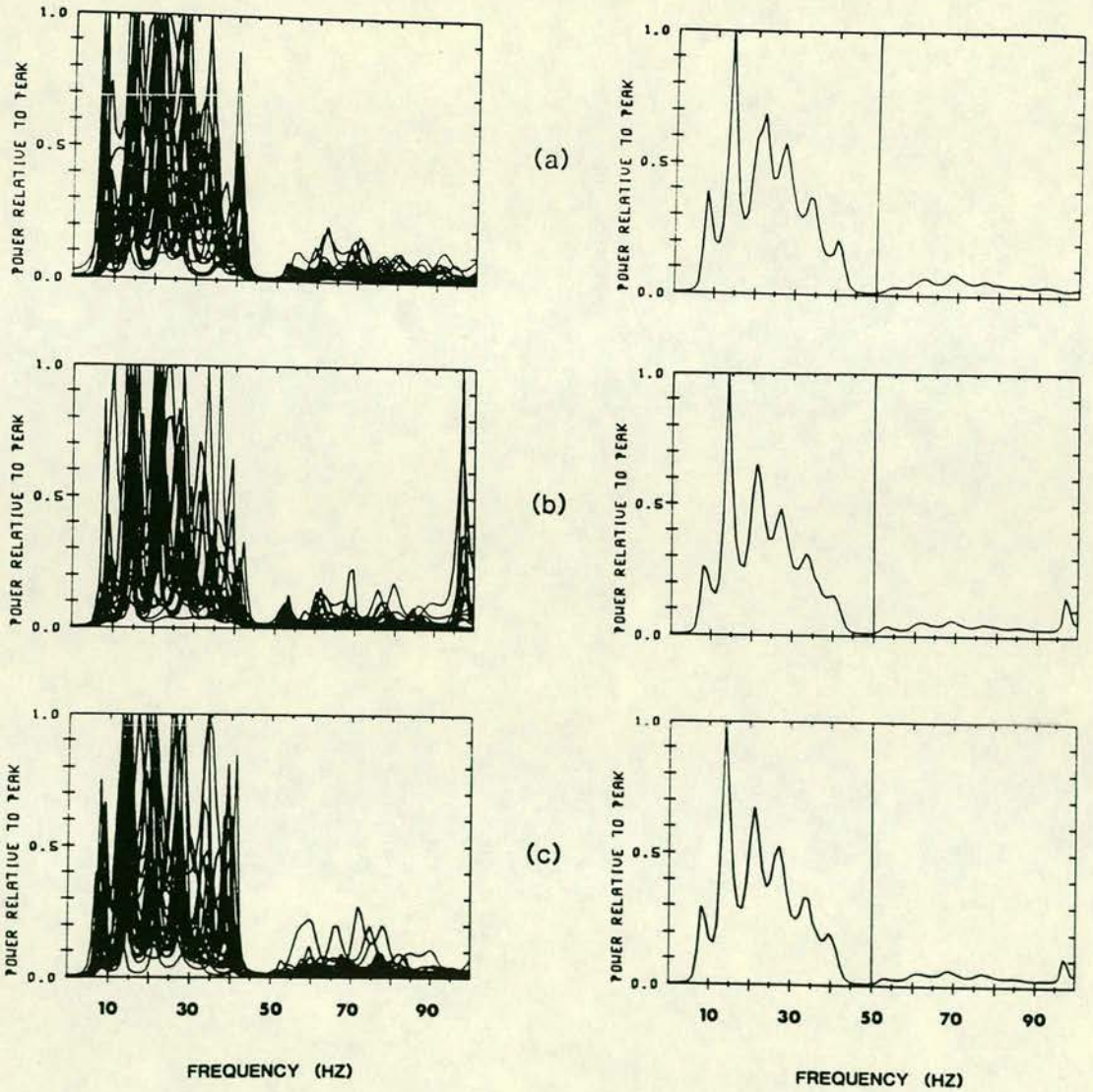


Fig. 6. Individual (left) and average (right) spectra obtained from successive data accumulations of 25 data windows. Decade 1 data, $M = 25$. Average spectra accumulate over all data windows. (a) Windows 1–25; (b) 26–50; (c) 51–75; (d) 76–100; (e) 101–125; (f) 126–150.

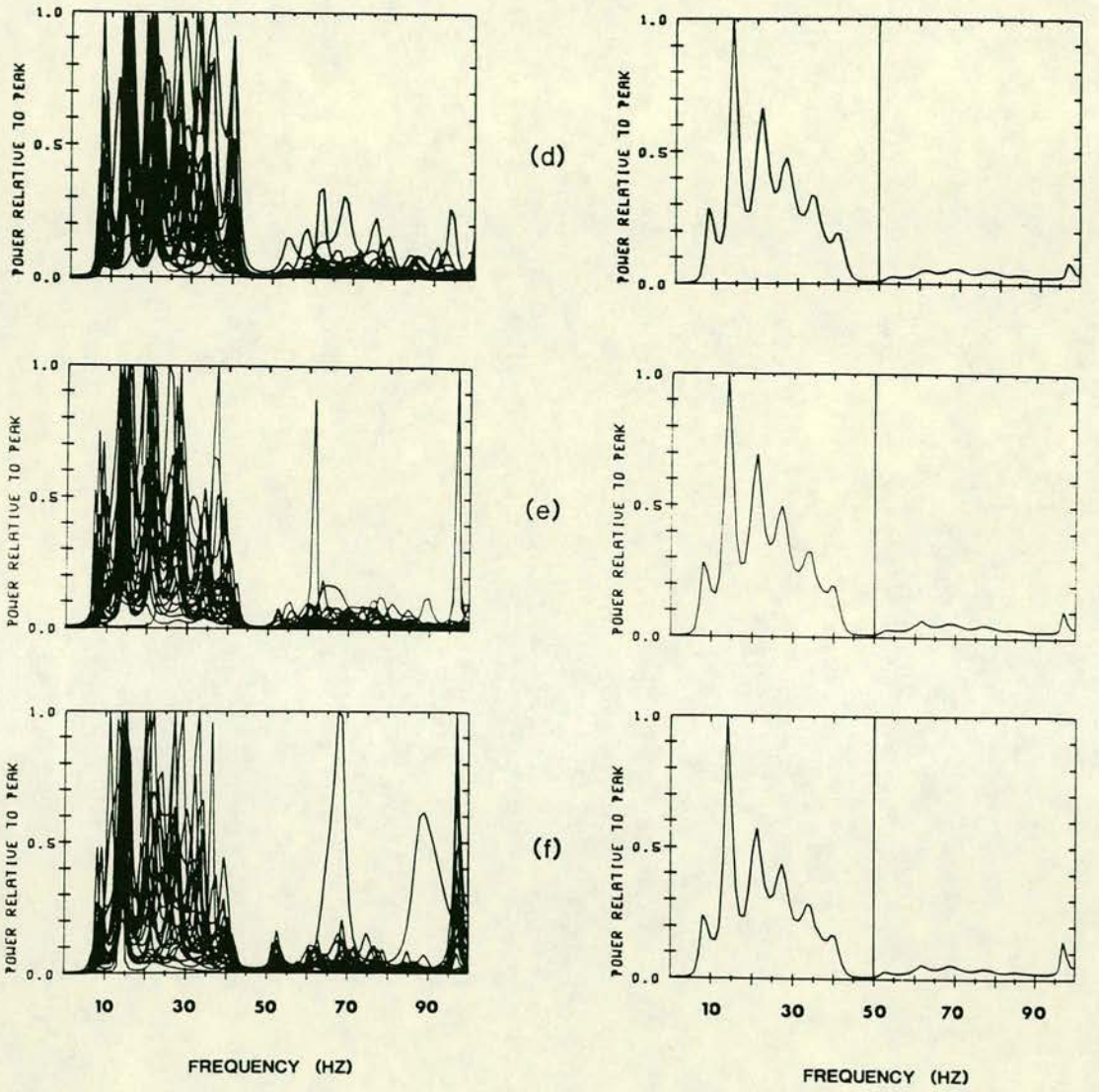


Fig. 6 continued.

Table 2. Resonance peak frequencies [f(Hz)] and the relative power levels (P) of the $n = 2$–6 Schumann resonance modes displayed in Fig. 6a–f

	f_2	P_2	f_3	P_3	f_4	P_4	f_5	P_5	f_6	P_6
(a)	14.0	1.00	22.0	0.68	27.0	0.58	33.5	0.38	40.0	0.21
(b)	14.0	1.00	21.0	0.66	27.0	0.49	33.0	0.31	39.5	0.15
(c)	14.0	1.00	21.0	0.68	27.0	0.53	33.5	0.33	39.0	0.20
(d)	14.0	1.00	21.0	0.67	27.0	0.48	33.0	0.34	39.5	0.19
(e)	14.0	1.00	21.0	0.70	27.0	0.50	33.5	0.33	39.0	0.20
(f)	14.0	1.00	21.0	0.58	27.0	0.44	33.0	0.29	39.0	0.17

mains frequency and its odd harmonics, together with weaker spectral lines at subharmonic frequencies (e.g. 100 Hz). In addition, the switching of current loads can produce broad-band transients and further high frequency harmonics. The 50 Hz notch filter mentioned previously has a centre frequency of 47.5 Hz and the attenuation achieved about this frequency is reflected in the spectra displayed in Fig. 6. A high frequency peak appears intermittently at around 97 Hz and is believed to be a power-line subharmonic. Other isolated large amplitude spectral components are occasionally detected at frequencies above 50 Hz, as shown in Figs. 6e and 6f. These are believed to be due to power-line transients.

There is an indication in the averaged spectral accumulations on the right of Fig. 6 that three higher order Schumann resonance modes ($n = 9, 10$ and 11) are detected at frequencies around 53, 62 and 69 Hz. The latter two modes were in fact the original experimental signatures of the Schumann resonances detected by BALSER and WAGNER (1960) and which were considered 'not statistically significant' at the time. The time local properties of the resonance frequencies are displayed as the overlaid plots on the left of Fig. 6. These plots provide some indication of the spectral variability encountered when individual waveforms are considered. The detailed nature of these spectra is considered later, following a discussion of the sources of cavity excitation.

A similar exercise was conducted for decade 2 data in order to establish the characteristics of the first order Schumann resonance peak. Spectral plots, equivalent to those of Fig. 6, are shown in Fig. 7 for a lower order ($M = 12$) filter length. This relatively low order of filter

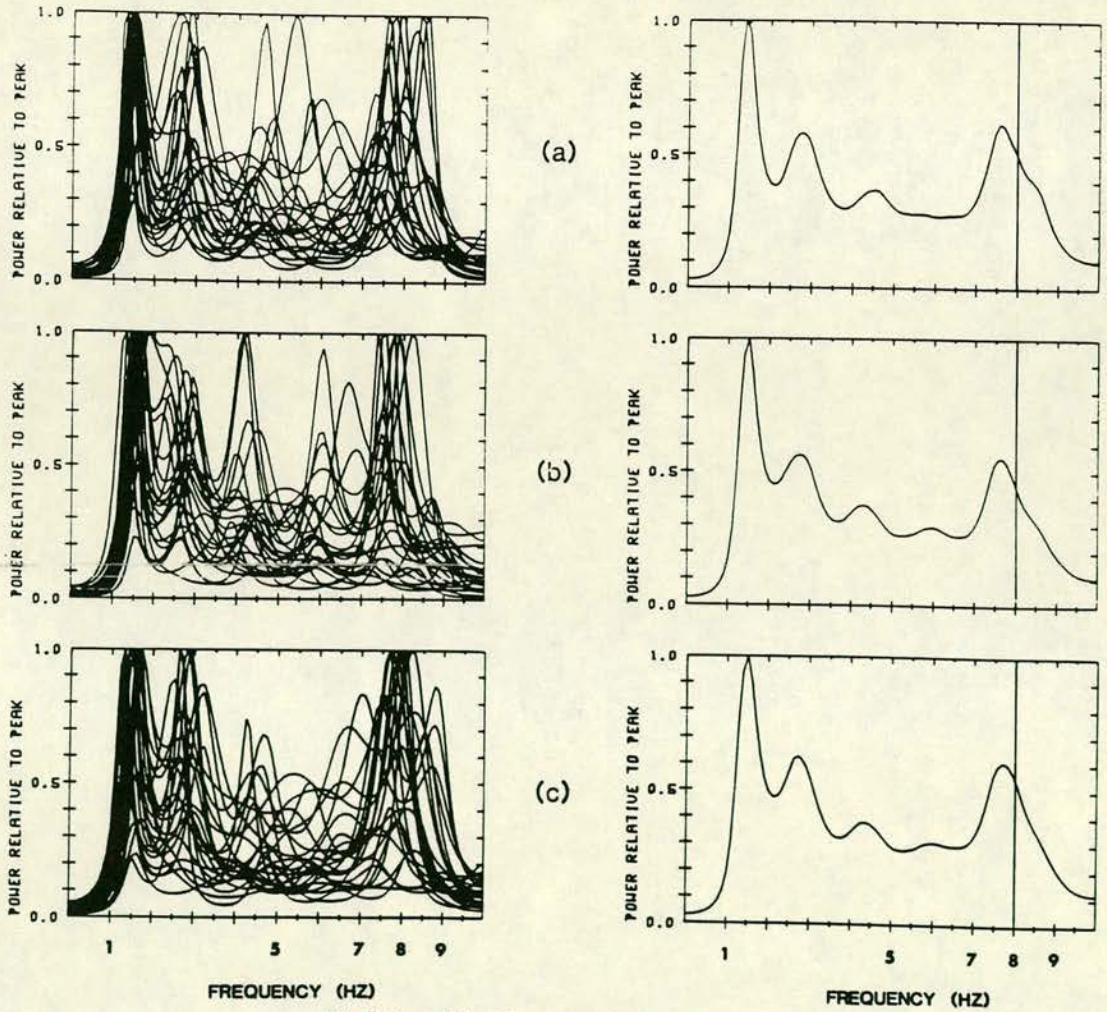


Fig. 7. As per Fig. 6, but decade 2 data and $M = 12$.

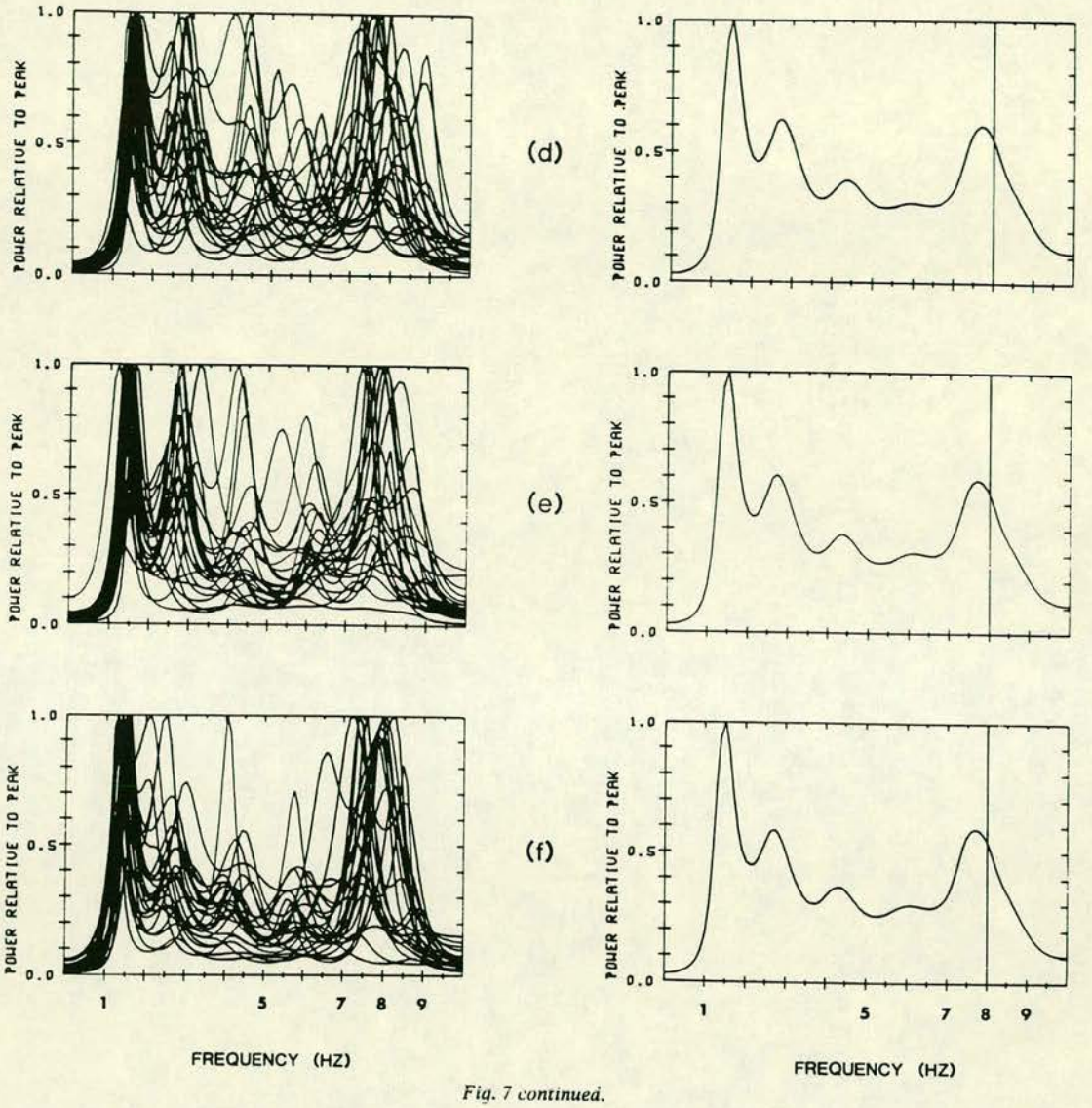


Fig. 7 continued.

length was chosen to provide resolution of a single first order resonance peak for comparison with previous experimental determinations. Each of the six sets of 25 data windows corresponds to a real time interval of 262.5 s. Despite the variability displayed in the overlaid plots of individual spectra, a stable resonance peak is obtained from the average spectra that varies in location between 7.50 and 7.65 Hz, over the six accumulated sets of data windows. These figures for the first order mode appear slightly lower than typical published values of about 7.8 Hz (MADDEN and THOMSON, 1965; RYCROFT, 1965). However, as we have already noted in Fig. 4a, the single peak resolved for $M = 12$ actually consists of two stable resonance peaks which can be distinguished by using higher order filter lengths. An equivalent plot for the first three sets of 25 data windows, obtained using $M = 18$, is shown in Fig. 8. The two peaks are established at frequencies of (a) 7.3 and 8.0, (b) 7.25 and 8.1, (c) 7.25 and 8.1 Hz. However, as can be seen in the overlaid data window plots, on the left of Fig. 8, individual spectra contain examples of both single and doublet Schumann resonance peaks. Such individual behaviour produces the varying contributions to the two resonance peaks defined in the average spectra of Figs. 8a, b and c. Such dynamic properties are better displayed using other methods of presentation, discussed below.

The above results confirm that for the data set

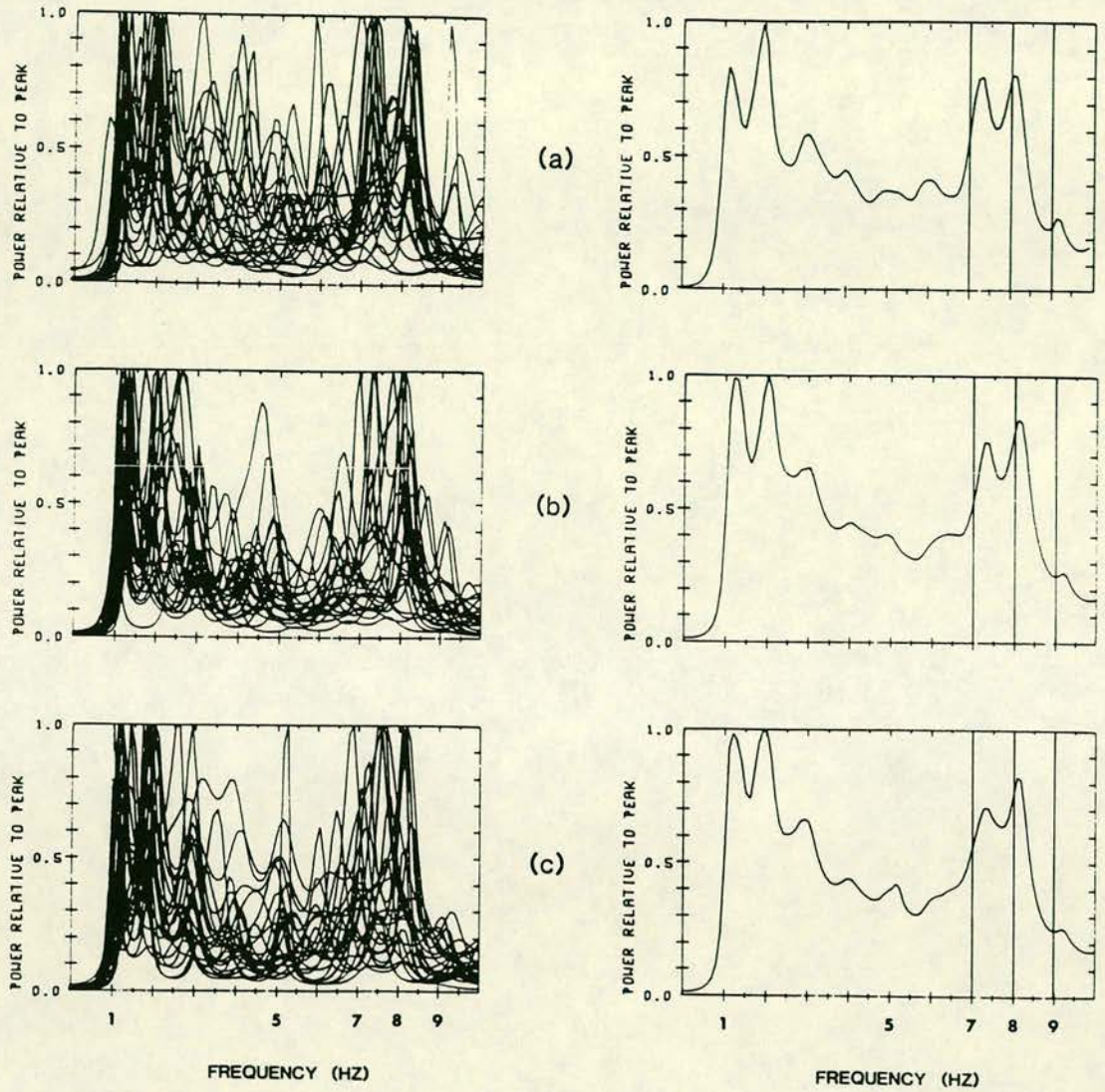


Fig. 8. As per Fig. 6, but decade 2 data and $M = 18$. (a) Windows 1–25; (b) 26–50; (c) 51–75.

considered, the ME technique provides reliable time averaged estimates of the low order ($n = 1$–6) Schumann resonance modes. There is also an indication in Fig. 6 that the higher order Schumann resonance modes are detectable at this moderate latitude (i.e. 40°N). In our opinion, however, the main advantage of the technique lies in the spectral resolution of individual waveforms which, according to the results presented, display complex frequency characteristics as a function of time. Many of these frequency characteristics are predicted by theory, as discussed below.

5. DISCUSSION

The digital AMT data considered in this study provide a bandwidth appropriate to studies of the Schumann resonance phenomena. The individual time windows are of the same order as the duration of many ELF events (e.g. Fig. 2). The study has demonstrated the possibilities that exist in the application of ME spectral techniques to such data. Using such techniques it appears that the spectral characteristics of individual Schumann resonance waveforms can be adequately resolved.

In the case of steady-state descriptions of the Earth–ionosphere waveguide properties, time-averaged characteristics may be required. The average spectra accumulated over time intervals of 94 s (decade 1) and 262.5 s (decade 2) shown in this study are examples of such properties. Although it is not our purpose here to investigate ionospheric parameters, the time averaged spectra of Figs. 6–8 would provide a basis for the evaluation of conductivity profiles according to the theory of TRAN and POLK (1979b). Within such accumulations time local variability is apparent, particularly in the case of the first order ($n = 1$) Schumann resonance mode. It has been demonstrated that the data for this mode contain time local frequency splitting. The waveguide appears to support the doublet frequencies of 7.3 and 8.1 Hz. A sonogram displaying the largest resonance peaks for the first mode across approximately 6 min is shown in Fig. 9a. Because of the highly resolved nature of the spectra, a non-linear contour interval is required. The resolution inherent in Fig. 9a reveals the dynamic nature of the first order Schumann resonance mode. The largest resonance peak detected is indeed a doublet, however, well defined singlet peaks occur (e.g. RUN 42) and a possible triplet resonance is observed around RUN 29. The dynamic nature of the characteristics observed in Fig. 9a indicates that the time averaged properties of this mode must be treated with caution.

An equivalent sonogram displaying the largest resonance peaks for the $n = 2$, 3, 4 and 5 and 6 Schumann resonance modes is shown in Fig. 9b. The time interval is restricted to approximately 3 min and again a non-linear contour interval has been used. The largest resonance peak occurs for the $n = 2$ mode and significant deviations from the average frequency of 14.0 Hz are observed. The same behaviour can be observed in the higher ($n = 3, 6$) modes as well. There is also an indication (e.g. RUN 27) that the $n = 3$ mode may comprise a doublet.

In order to understand the results presented, it is first necessary to consider the nature of the waveforms observed (e.g. Fig. 1), which are responsible for the spectra displayed in Fig. 9. From Fig. 1 we can distinguish two of the three possible signal components

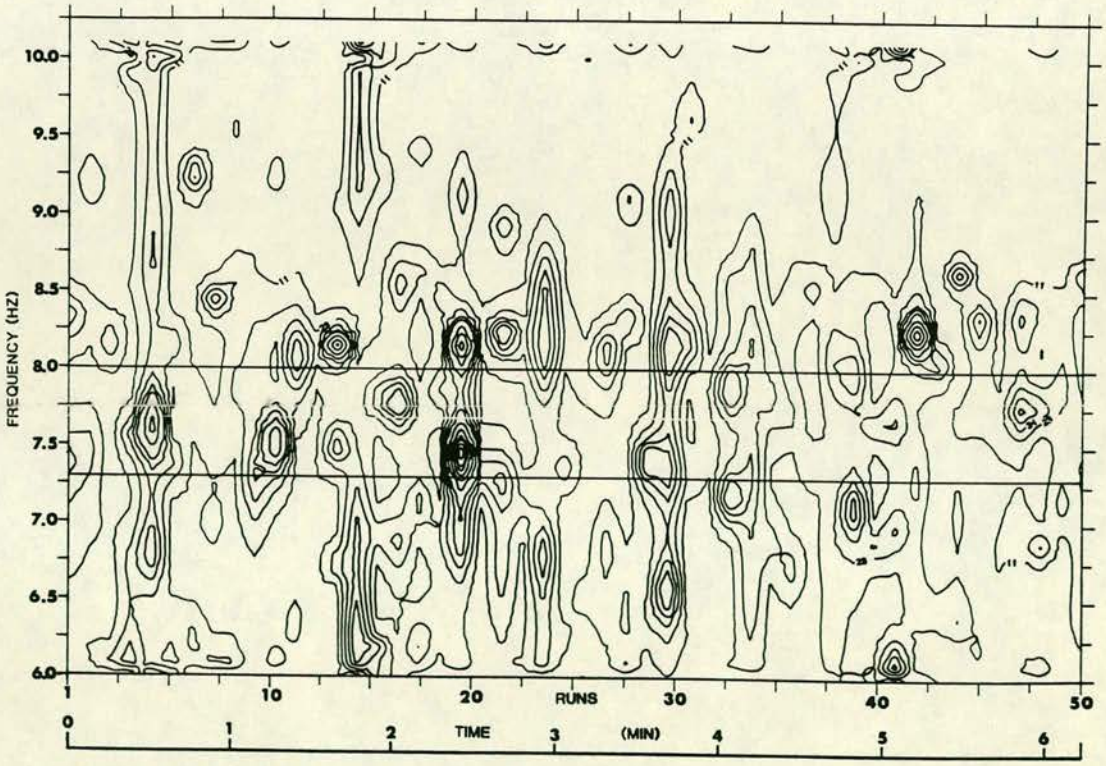


Fig. 9. High resolution sonograms of the Schumann resonance modes. Power density values in $(nT)^2\,Hz^{-1} \times 10^{-8}$. Non-linear contour interval with minimum value of 0.1×10^{-10}. (a) $n = 1$ mode; decade 2 data; maximum value 0.23×10^{-5}; filter length $M = 25$. (b) $n = 2$, 3, 4, 5 and 6 modes; decade 1 data; maximum value 0.15×10^{-5}; filter length $M = 35$.

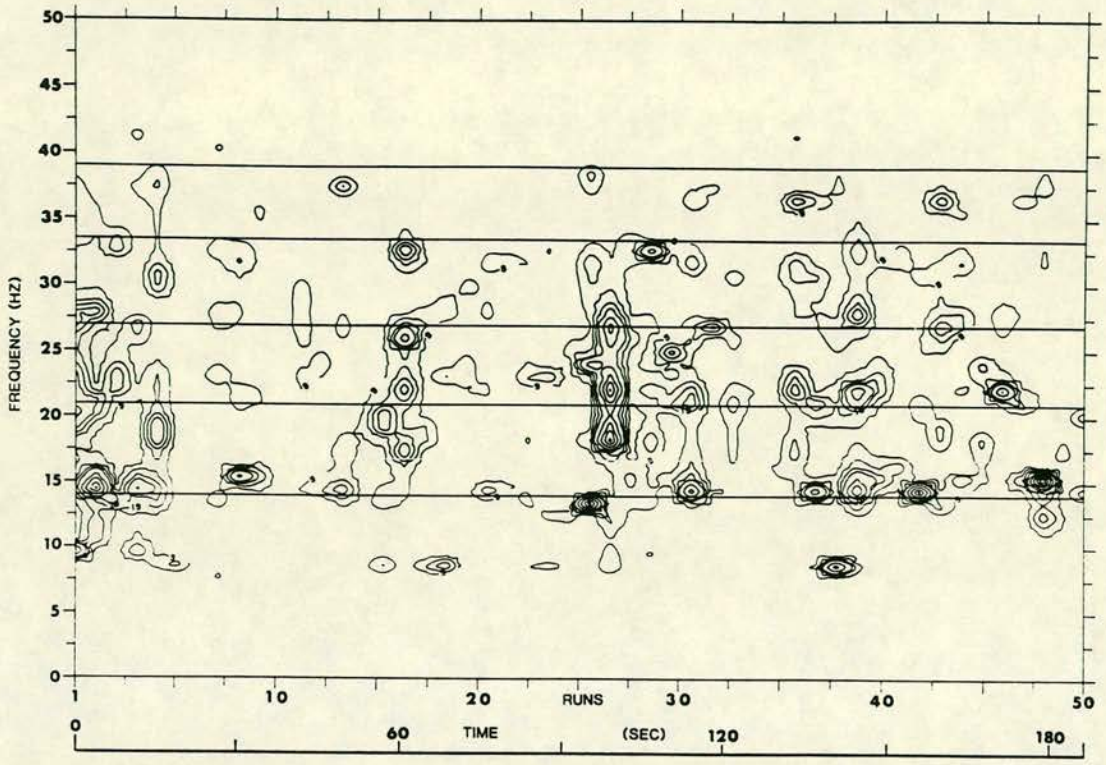


Fig. 9 continued.

considered by BLIOKH *et al.* (1980, p. 121) to comprise distinct sources of cavity excitation. The first component is the Schumann background, which is always present and which is the response of the cavity to continuous global lightning activity. The second category of signal component is referred to as an ELF transient event and typically exceeds the background by a factor of 2–10. According to JONES and KEMP (1971), the forms of such transient events may vary between a prolonged burst of complex field variations (e.g. Fig. 1, RUN 4) to a simple damped quasi-sinusoidal oscillation (e.g. Fig. 2). The two consecutive waveforms displayed in Fig. 2 possess the characteristic commencement of typical transient events noted by JONES and KEMP (1971, fig. 5). JONES and KEMP (1970, 1971) undertook both theoretical and experimental studies on such ELF events. In conclusion the authors suggest that such events are due to the transient excitation of cavity resonances by unusually large lightning discharges, the majority of which are cloud discharges.

The sonograms of Fig. 9 display the spectra associated with individual data runs. Obviously the resonances possessing the largest power densities arise from data runs containing larger than average transient events superimposed on the background. The lowest power densities arise from the continuum of background sources. The sonograms of Fig. 9 necessarily emphasise the large amplitude transient events. We first consider the resonance characteristics for the $n = 2, 3, 4, 5$ and 6 modes displayed in Fig. 9b. According to JONES and KEMP (1970), the most important characteristic of the Schumann resonance spectrum is that, for a given source and fixed ionosphere, the spectral structure is a unique function of propagation distance. The theoretical computations on spectral behaviour as a function of distance from the source (JONES and KEMP, 1970, 1971) indicate an overall spectral structure that is strongly dependent on distance. We attribute much of the variability in resonance behaviour displayed in Fig. 9b to such an effect. The main reason for this conclusion stems from the variety of horizontal magnetic field polarisation vectors displayed by the transient events. It is hoped to report on this work at a later date.

Turning now to the fine structure displayed by the first order mode in Fig. 9a. It is apparent that over relatively short periods of time (i.e. tens of seconds), the

first order resonance mode is characterised by the presence of singlet, doublet and possibly triplet resonance peaks. We suggest that, since such characteristics are a pervasive feature of our data, they constitute both the split and non-split first order resonance modes predicted by theory and discussed in the introduction. We therefore attribute the observed fine structure to properties inherent in the cavity system, rather than to source-dependent characteristics. Such a conclusion requires that the mechanisms responsible for degeneracy within the cavity are highly dynamic over time scales of the order of minutes.

This initial empirical study is not exhaustive with regard to the available data. A small portion of the data set has been used to illustrate the use of a data adaptive spectral technique on electromagnetic data obtained at audio frequencies. The spectra resolution afforded by the technique is revealing and further work on the spectral and polarization properties of the waveforms is in progress.

Acknowledgement—The authors wish to thank a referee for useful advice and constructive comments. This paper is published with the approval of the Director, British Geological Survey (NERC).

REFERENCES

Akaike H.	1969	*Ann. Inst. statist. Math., Tokyo* **21**, 407.
Akaike H.	1970	*Ann. Inst. statist. Math., Tokyo* **22**, 203.
Akaike H.	1974	*IEEE Trans. autom. Control* **AC-19**, 716.
Akaike H.	1976	*System Identification—Advances and Case Studies* (Mehra R. K. and Lainiotis D. G., Eds). Academic Press, New York.
Balser M. and Wagner C. A.	1960	*J. Res. natn. Bur. Stand.* **64D**, 415.
Berryman J. G.	1978	*Geophysics* **43**, 1984.
Bliokh H., Nikolaenko A. P. and Filippov Yu. F.	1980	*Schumann Resonances in the Earth–Ionosphere Cavity* (Llanwyn-Jones D., Ed.), IEE Electromagnetic Wave Series 9. P. Peregrinus Ltd, Stevenage.
Burg J. P.	1968	Paper presented at Advanced Study Institute on Signal Processing, NATO, Enschede, Netherlands.
Cannon P. S. and Rycroft M. J.	1982	*J. atmos. terr. Phys.* **44**, 201.
Fischer G.	1982	*Geophys. Surv.* **4**, 373.
Fougere P. F.	1977	*J. geophys. Res.* **82**, 1051.
Galejs J.	1961	*IRE Trans. Antennas Propagn* **AP-9**, 554.
Galejs J.	1964	*Natural Electromagnetic Phenomena Below 30 kc/s*, (Bleil D. F., Ed.). Plenum Press, New York.
Galejs J.	1972	*Terrestrial Propagation of Long Electromagnetic Waves*. Pergamon Press, New York.
Jaynes E.	1982	*Proc. IEEE* **70**, 939.
Jenkins G. M. and Watts D. G.	1968	*Spectral Analysis*. Holden-Day, San Franscisco.
Jones D. L. and Kemp D. T.	1970	*J. atmos. terr. Phys.* **32**, 1095.
Jones D. L. and Kemp D. T.	1971	*J. atmos. terr. Phys.* **33**, 557.
Keller G. V.	1971	*Geoexploration* **9**, 99.
Levinson H.	1947	*J. Math. Phys.* **25**, 261.
Madden T. and Thomson W.	1965	*Rev. Geophys.* **3**, 211.
Parzen E.	1976	Technical Report 37, Statistical Science Division, State University of New York in Buffalo.
Polk C.	1969	*Planetary Electrodynamics*, Vol. 2 (Coroniti S. C. and Hughes J., Eds). Gordon and Breach, New York.
Rycroft M. J.	1965	*J. Res. natn. Bur. Stand.* **69D**, 1071.
Schumann W. O.	1952	*Z. Naturf.* **7A**, 149.
Sentman D. D.	1983	*J. atmos. terr. Phys.* **45**, 55.
Strangway D. W., Swift S. M. Jr and Holmer R. C.	1973	*Geophysics* **38**, 1159.
Tran A. and Polk C.	1976	*Radio Sci.* **11**, 803.
Tran A. and Polk C.	1979a	*J. atmos. terr. Physl* **41**, 1241.
Tran A. and Polk C.	1979b	*J. atmos. terr. Phys.* **41**, 1249.
Ulrych T. J. and Bishop T. N.	1975	*Rev. Geophys. Space Phys.* **13**, 183.
Ulrych T. J. and Clayton R. W.	1976	*Phys. Earth Planet. Interiors* **12**, 188.

Journal of Atmospheric and Terrestrial Physics, Vol. 49, No. 3, pp. 217–229, 1987.
Printed in Great Britain.

0021–9169/87 $3.00 + .00
Pergamon Journals Ltd.

Time domain polarization analysis of Schumann resonance waveforms

A. TZANIS* and D. BEAMISH

British Geological Survey, Murchison House, West Mains Road, Edinburgh EH9 3LA, U.K.

(*Received in final form* 22 *April* 1986)

Abstract—Schumann resonance waveforms in the lower ELF band are produced by distant lightning discharges. The multiplicity of world-wide thunderstorm centres provides a background spectrum which is generally incoherent. Superimposed on this background are larger amplitude events from individual thunderstorm centres. Source direction finding requires the determination of the polarization properties of these complex and superimposed waveforms. A time domain polarization analysis technique is described and applied to a variety of examples of Schumann resonance waveforms. The background waveforms display the anticipated level of incoherent polarization properties. It is demonstrated that the technique can isolate individual waveforms that are linearly polarized and which represent the arrival of short path signals from single source locations.

INTRODUCTION

In the ELF band (5 Hz–25 kHz) electromagnetic energy is largely produced as radiation from vertical lightning discharges. It is estimated that of the order of 100 lightning flashes occur each second (TURMAN and EDGAR, 1982). Such sources may contain a wide spectrum of frequencies, however, for distant sources the energy undergoes multiple reflection in the Earth–ionosphere cavity giving rise to frequency dependent attenuation. The main energy components in the lower ELF band (5 Hz–2 kHz) are the Schumann resonance modes. The energy propagates as a transverse magnetic mode in which the electric field is largely radial and the magnetic field horizontal (MADDEN and THOMSON, 1965). Although the polarization characteristics of the horizontal magnetic field are required to successfully locate a discharge centre (KEMP, 1971; BLIOKH *et al.*, 1980), only a few examples of the polarization properties of ELF waveforms have appeared in the literature.

The data used in the present analysis have been previously examined in the high-resolution spectral study of BEAMISH and TZANIS (1986). The AudioMagnetoTelluric (AMT) instrumental conditioning and collection scheme provides four decades of data from 100 to 0.01 Hz. Decade 1 data are sampled at 200 Hz and provide a data bandwidth from 100 to 10 Hz. This decade attenuates the fundamental (7.5 Hz) Schumann resonance mode but provides data for studies of the higher order modes. It is data from this decade, part of the lower ELF band, that is the subject of the present study. The data were collected in western Anatolia (Turkey); geographic coordinates 40.5°N, 30.0°E.

In general, Schumann resonance waveforms comprise 'events' superimposed on a background. The spectral content of typical decade 1 data has been described by BEAMISH and TZANIS (1986). Stable Schumann resonance modes are consistently resolved at 14.0, 21.0, 27.0, 33.5 and 39.5 Hz by integrating the spectrum over real time windows of approximately 90 s. An examination of the time local spectral properties of Schumann waveforms revealed large variability in power spectral densities and their frequency content. The sonogram of Fig. 1 is used to illustrate this behaviour. The lowest power levels arise from the continuum of background activity. The peak power levels are due to large amplitude transient events. JONES and KEMP (1970, 1971) indicate that for a given ionosphere the spectral structure of Schumann waveforms is strongly dependent on distance from the source. BEAMISH and TZANIS (1986) attributed the spectral variability indicated in Fig. 1 to the effect of signals being received from a variety of distributed source locations. The present polarization analysis of the horizontal magnetic field enables this assertion to be examined in detail.

In order to extract realistic polarization parameters for the waveforms considered it has been necessary to investigate suitable time local polarization techniques.

*Also at: Geophysics Department, University of Edinburgh, Kings Buildings, J.C.M.B., Mayfield Road, Edinburgh, U.K.

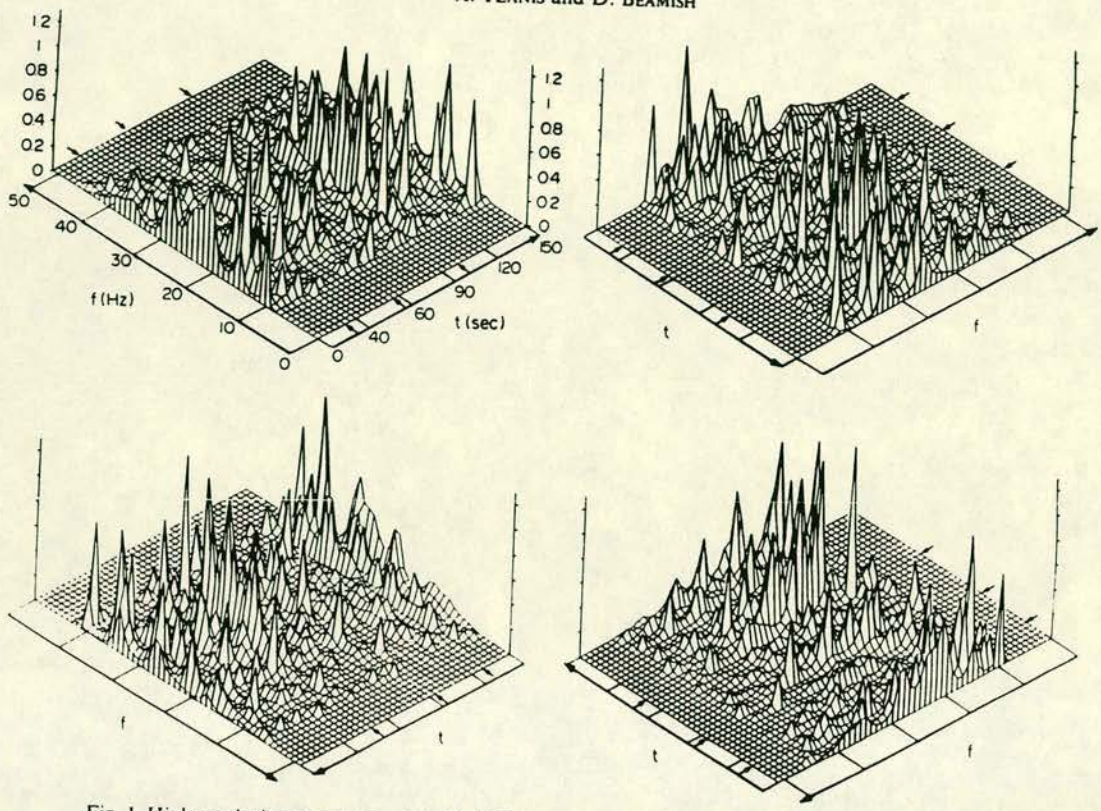


Fig. 1. High resolution sonogram of the N–S (H) magnetic component. Four perspective views of the same sonogram. Data collected on 8 June 1984. Power density values in $(nT)^2\ Hz^{-1} \times 10^{-6}$; $n = 2, 3, 4, 5, 6$ modes; decade 1 data; Maximum Entropy power spectra (see BEAMISH and TZANIS, 1986); filter length $M = 35$.

A monochromatic, elliptically-polarized wave of angular frequency w may be defined by its two orthogonal components

$$x(t) = X(t) \cdot \cos\left[wt + \phi_x(t)\right],$$

$$y(t) = Y(t) \cdot \cos\left[wt + \phi_y(t)\right],$$

e.g. the two horizontal magnetic components (H, D). The basic theory of polarization analysis is then as given by BORN and WOLF (1959) and repeated by FOWLER *et al.* (1967). The polarization characteristics are usually derived by examining the coherence or cross-spectral matrix between the two components. Such characteristics are obtained in the *frequency* domain across appropriate time averages. In the present study we adopt a technique which permits the derivation of polarization parameters in the time domain, in the form of instantaneous values. The advantage of this approach is that it permits the tracing of any variation in the state of *individual* waveforms encountered within a particular data window.

DATA ANALYSIS AND THE CONCEPT OF A COMPLEX SIGNAL

KODERA *et al.* (1977) describe a method for the analysis of polarization of any two orthogonal components (x, y) based on the complex representation of a polarized signal. The technique was first described by GABOR (1946). The technique extends the formalism obeyed by an *analytic* signal to a two component representation. The method proposed by KODERA *et al.* (1977) was primarily intended for measurements made in rotating frames of reference (satellites and spacecraft), but it is also suitable for a wide range of applications, such as narrow band or linearly polarized signals in fixed, although arbitrary, frames of reference. Our AMT data exhibit large waveform variability and are recorded in relatively narrow band schemes over time intervals comparable to those of many transient ELF events. In the present study we use the same mathematical formulation as KODERA *et al.* (1977), but a different numerical implementation. A detailed account of the formalism

can be found in KODERA *et al.* (1977) and only a brief description is provided here.

For a given monochromatic signal $x(t)$, defined by

$$x(t) = X(t)\cdot\cos[wt+\phi(t)],$$

a complex or analytical signal $x_a(t)$ may be defined by

$$x_a(t) = x(t)+i\tilde{x}(t),$$

where $\tilde{x}(t)$ is the Hilbert transform of $x(t)$. The instantaneous amplitude $X(t)$, the instantaneous phase $\phi(t)$ and the instantaneous frequency $f(t)$ are then given by

$$X(t) = [x^2(t)+\tilde{x}^2(t)]^{1/2},$$

$$\phi(t) = \tan^{-1}[\tilde{x}(t)/x(t)],$$

$$f(t) = \frac{d}{dt}[\phi(t)].$$

The complex signal associated with a real signal having two orthogonal components (x, y) is simply defined as

$$c(t) = x(t)+iy(t).$$

KODERA *et al.* (1977) demonstrate that it is possible to express the complex signal in terms of circular components rotating in opposite directions. At any given frequency w_0 we may write

$$c(t, w_0) = C(w_0)\cdot\exp(iw_0t),$$

which represents a circular harmonic in the complex plane rotating in a sense depending on the sign of w_0. Positive frequencies correspond to polarized signals rotating in the positive (counterclockwise) sense and negative frequencies to signals rotating in the negative (clockwise) sense. The power spectrum of the complex signal is related to the intensity of the circular components and the spectral asymmetry is linked to the polarization of the real signal.

It is possible to divide the frequency spectrum of the complex signal into two parts, corresponding to positive (+) and negative (−) frequencies, such that

$$C(w) = C+(w)+C-(w).$$

By inverse Fourier transformation we then obtain

$$c(t) = c+(t)+c-(t).$$

KODERA *et al.* (1977) show how these circular decomposites of the complex signal can be expressed in terms of the analytic signals for x and y. The expressions are

$$c+(t) = [x_a(t)+iy_a(t)]/2,$$

$$c-(t) = [x_a^*(t)+iy_a^*(t)]/2.$$

The polarization parameters are conventionally expressed in terms of the amplitudes and phases of the two orthogonal components. The above formulation, however, provides expressions in terms of the circular components of the signal. A circular signal is not affected by the orientation of the coordinate system and therefore the amplitude of the complex signal remains invariant and all the information about the polarization axis is contained in the phase term. We now go on to obtain the relationships between the conventional polarization parameters and the two circular components $c+$ and $c-$. We consider the case of monochromatic, elliptically-polarized signals only.

Consider two circular components as given above with

$$c+(t) = A_1\cdot\exp[i(wt+\phi_1)],$$

$$c-(t) = A_2\cdot\exp[i(wt+\phi_2)].$$

When the two components are in phase the signal reaches the maximum amplitude $(|A_1+A_2|)$, which represents the major axis of the polarization ellipse. When they are out of phase the signal reaches its minimum amplitude $(|A_1-A_2|)$, which represents the minor axis of the polarization ellipse. The polarization parameters are therefore simply obtained as

$$\text{half major axis: } a = |c+|+|c-|,$$

$$\text{half minor axis: } b = |c+|-|c-|,$$

$$\text{ellipticity: } E = b/a,$$

$$\text{azimuth: } \theta = [\arg(c+)+\arg(c-)]/2.$$

In these expressions b and E are positive for counterclockwise rotation, and vice versa. Although all the above expressions assume monochromatic waveforms, KODERA *et al.* (1977) suggest that these formulae may also be used to obtain instantaneous polarization parameters for non-monochromatic real signals. The numerical implementation of the procedure is described in detail by TZANIS and BEAMISH (1986).

TIME DOMAIN HODOGRAMS

Prior to the presentation of the polarization properties obtained using the above procedures, we first examine simple time domain hodograms. Throughout the results presentation we use 150 data points of decade 1 data with a sampling rate of 200 Hz. The components used are the two horizontal magnetic field components (H, D), expressed in nT. Time domain hodograms simply plot the locus of the rotation vector of the two orthogonal components in

the horizontal plane. For a given data window (150 points) we may plot the rotation vector over the whole window or over selected subsets. Here we display the two components over the whole data window and plot hodograms for the five successive subsets of 30 points encountered within each window. The hodograms of each subset are independently scaled. The procedure thus approximates the selection of individual waveforms.

The data were collected at a single site on 26 July 1984 commencing at 12.40 h (local time). The results from three data windows are shown in Fig. 2. The largest peak-to-peak amplitudes are all less than 0.05 nT, and where waveform amplitudes fall below 0.01 nT we begin to encounter noise. The RUN numbers refer to the collection number of each data window. Successive data windows are separated by approximately 3 s in real time. We present the results to convey the general complex nature of the sferic waveforms encountered. It can be seen that the waveforms provide the whole range of elliptical polarization properties from linear through to quasi-circular. It can also be seen that the time domain hodograms offer only a relatively crude method of presenting the polarization characteristics of these complex waveforms.

TIME DOMAIN POLARIZATION OF BACKGROUND SFERICS

Detailed polarization results are now presented for 3 data windows which we consider to be representative of background sferic source fields. The data were collected at a single site on 8 June 1984 commencing at 10.40 h (local time). The same method of presentation is repeated for all three examples. For each example shown in Fig. 3 the polarization parameters ellipticity (E) and azimuth (θ) are plotted in the time domain below the corresponding data. This method of display enables the variations in the state of polarization of individual waveforms to be monitored. The results obtained have been fully verified using the above simpler techniques. The instantaneous values of (E, θ) consist of a series of discrete symbols that correspond to the extrema in (H, D), i.e. the signal inflexion points evaluated using the modulus. If the modulus falls below 20% of the maximum excursion encountered the result is omitted. This method of display is a simple and effective method of suppressing the contributions from low amplitude waveforms. In keeping with the previous definitions, a value of $E = 1$ corresponds to a circular wave rotating counterclockwise, a value of $E = 0$ corresponds to a linearly polarized wave and a value of $E = -1$ to a circular wave rotating in a

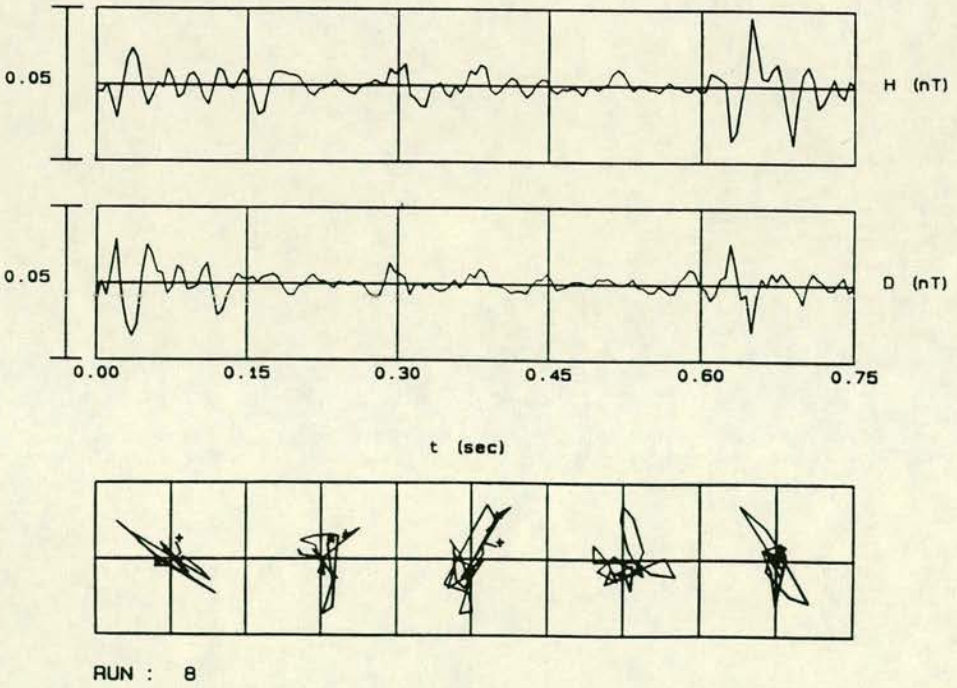


Fig. 2. Time domain hodograms. H and D correspond to the magnetic N-S and E-W components.
(*Continued over.*)

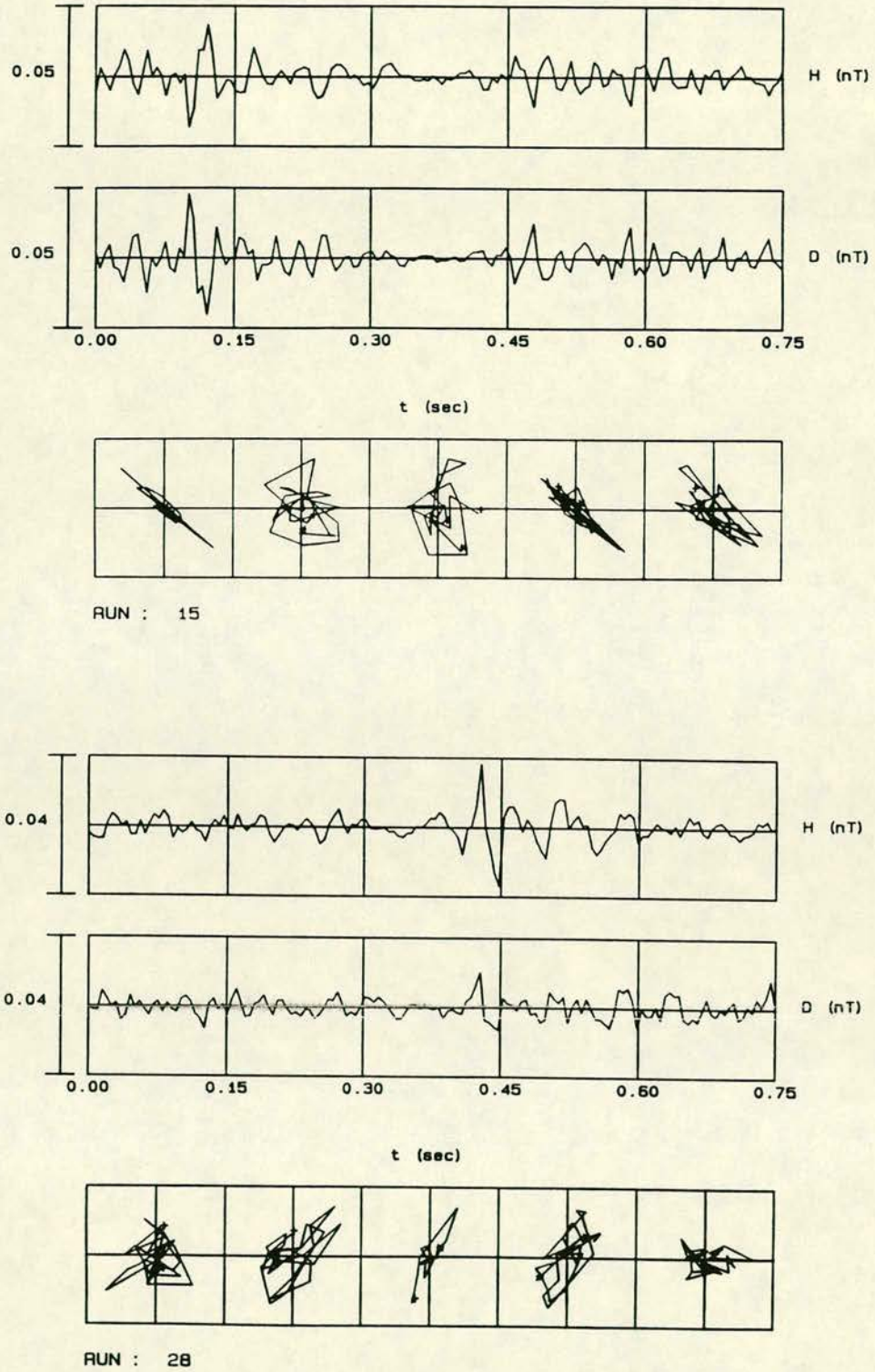


Fig. 2 continued.

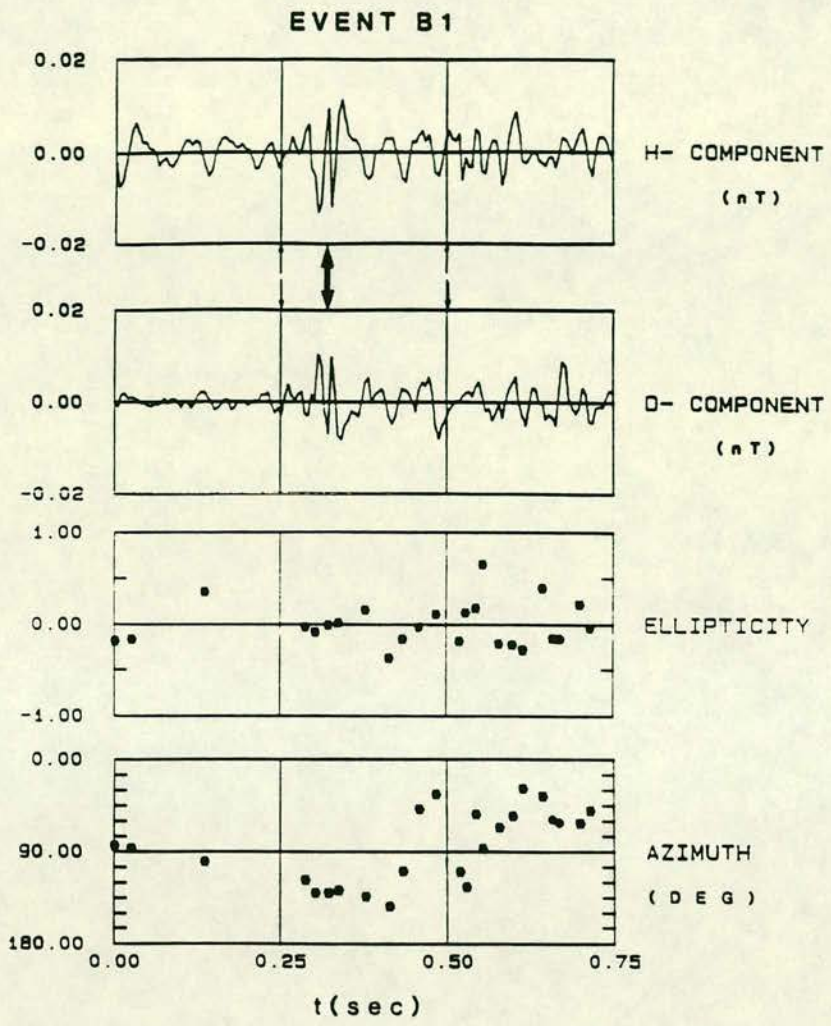


Fig. 3. Three examples of the time domain polarization analysis of background sferics. *H* and *D* correspond to the N–S and E–W magnetic components; real time separation of events B1–B2, 37.5 s, events B2–B3, 45.00 s. (*Continued over.*)

clockwise sense. Azimuths are defined in magnetic coordinates of west (180°), north (90°) and east (0°).

It can be seen from the examples provided in Fig. 3 that the largest peak-to-peak amplitude is of the order of 0.03 nT. The peak-to-peak amplitude of the general background appears to be less than 0.01 nT for this particular day. The full range of ellipticities and azimuths appear to be sampled, although none of the waveforms could be considered to be circularly polarized. According to BLIOKH *et al.* (1980), the background waveforms are the response of the cavity to the nearly continuous global sum of lightning activity. The waveforms received are thus the superposition of arrivals from multiple and incoherent sources. If this is the case, the polarization characteristics of background waveforms should be effectively random. Such characteristics are observed most acutely in many of the waveforms of Fig. 3b.

When we consider several of the larger amplitude events shown in Fig. 3 (indicated by thick arrows) we note a tendency for the waveforms to be linearly polarized at a given azimuth. The azimuths are different in each case and the polarization characteristics remain stable over several cycles. It therefore appears that the present method of analysis is capable of detecting 'background level' resonance waveforms from individual source locations. The data windows shown above belong to the same ensemble that was used to construct the sonogram of Fig. 1, and their approximate position is indicated with arrows.

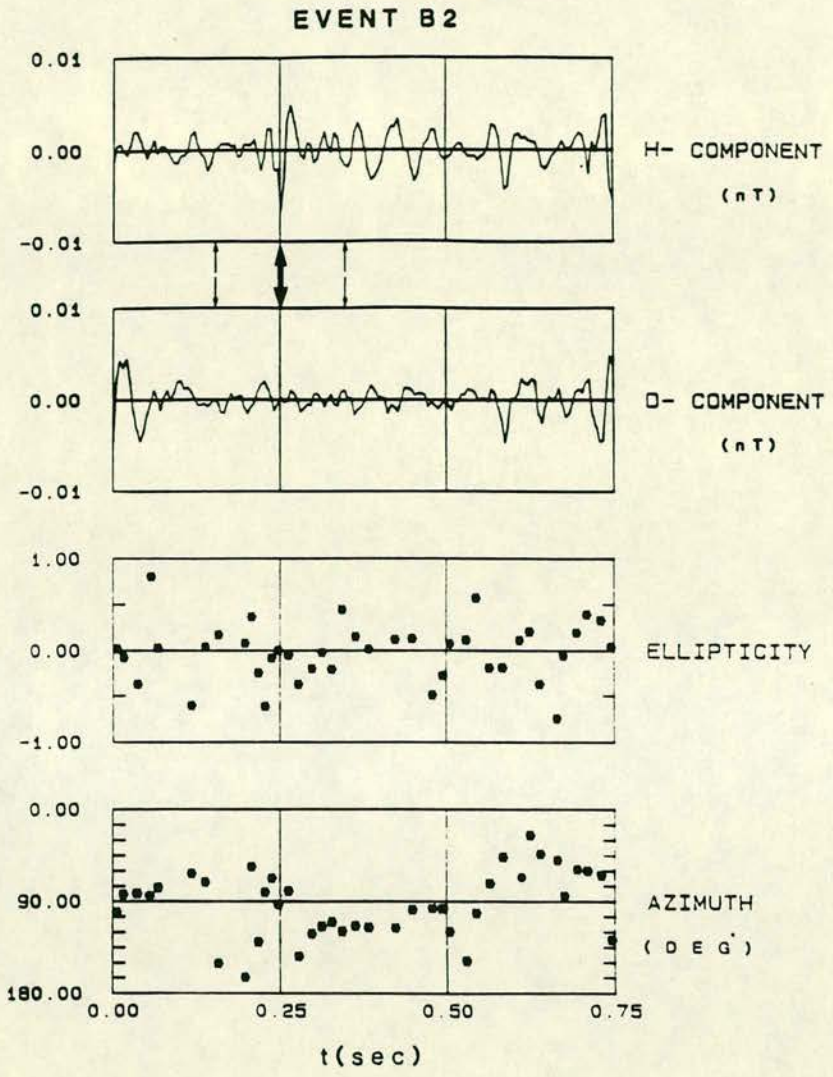


Fig. 3 continued. (*Continued over.*)

TIME DOMAIN POLARIZATION OF IMPULSIVE SFERICS

Detailed polarization results are now presented for three data windows which we consider to be representative of larger amplitude, impulsive ELF fields. According to JONES and KEMP (1971), such events are due to the transient excitation of the cavity resonance by unusually large lightning discharges. The data were collected at a single site on 27 July 1986 commencing at 10.37 h (local time). The three examples of impulsive events shown in Fig. 4 all have peak-to-peak amplitudes greater than 0.03 nT, the largest being of the order of 0.1 nT (Fig. 4b). The impulsive waveforms can all be characterized as damped, quasi-sinusoids. The three examples all possess the characteristic commencement of typical transient events noted by JONES and KEMP (1971). The initial part of each waveform is due to the arrival of the direct (short path) signal. Subsequent fluctuations are due to the arrival of the long path signal and to the circulation of the energy around the globe. It is quite evident from the polarization results that these waveforms are linearly polarized at a particular azimuth and the azimuths differ for each of the three examples. The two main waveforms encountered in Fig. 4b provide a good example of single source repetition on a time scale of 0.5 s.

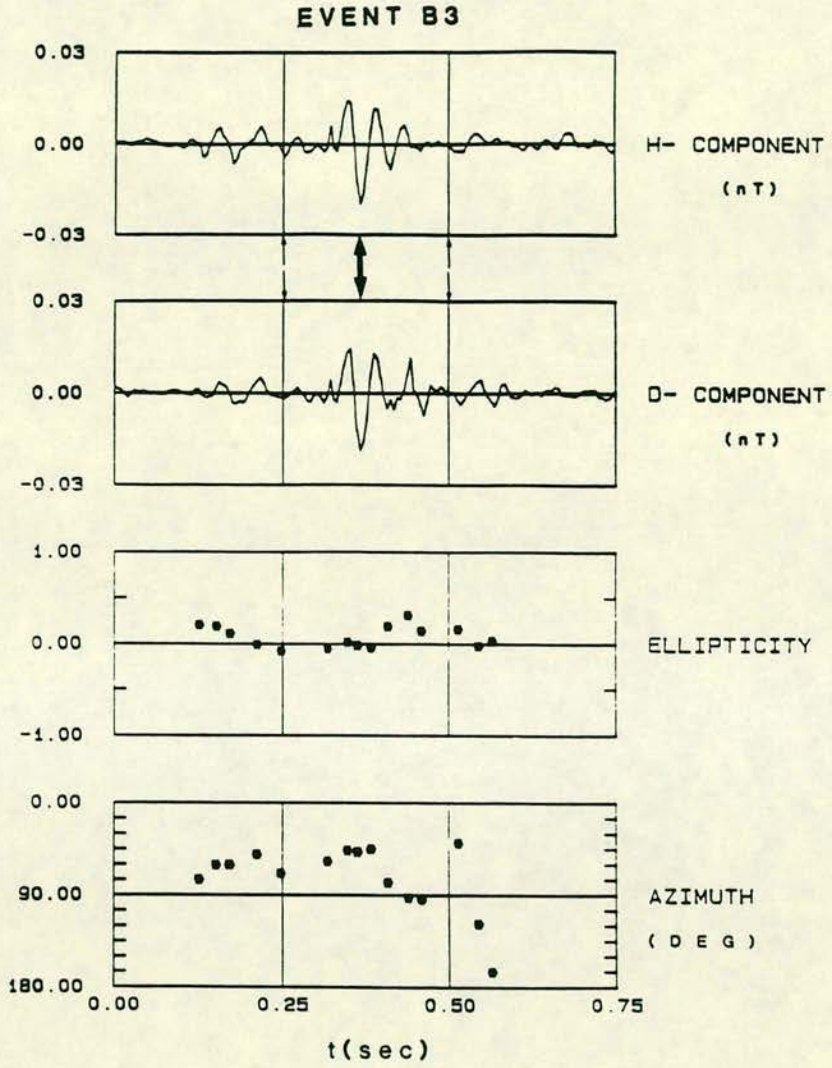


Fig. 3 continued.

SPECTRAL SIGNATURES OF THE EXAMPLE SFERICS

The polarization analysis of the background and impulsive Schumann resonance waveforms reveals that they contain a variety of events with a wide range of polarization vectors, which indicates the existence of distributed sources and simultaneous thunderstorm activity.

The theoretical and experimental work of JONES and KEMP (1970, 1971) and KEMP (1971) suggests that the spectral shape, i.e. the relative position of peaks and troughs, is a function of the source receiver separation for a specified ionosphere. Given the azimuthal distribution of the sources, it is anticipated that the spectral signature of our example waveforms will vary in a manner characteristic of the source spectrum and the propagation distance. This could account for the extremely variable time local behaviour of the Schumann resonances, as indicated in Fig. 1 (see also fig. 9 of BEAMISH and TZANIS, 1986). We now proceed to investigate this assertion.

In Fig. 5a–c we present the power spectra for the azimuthal (total horizontal) magnetic field of the example background events, annotated B1, B2 and B3, respectively, and in Fig. 5d–f the equivalent power

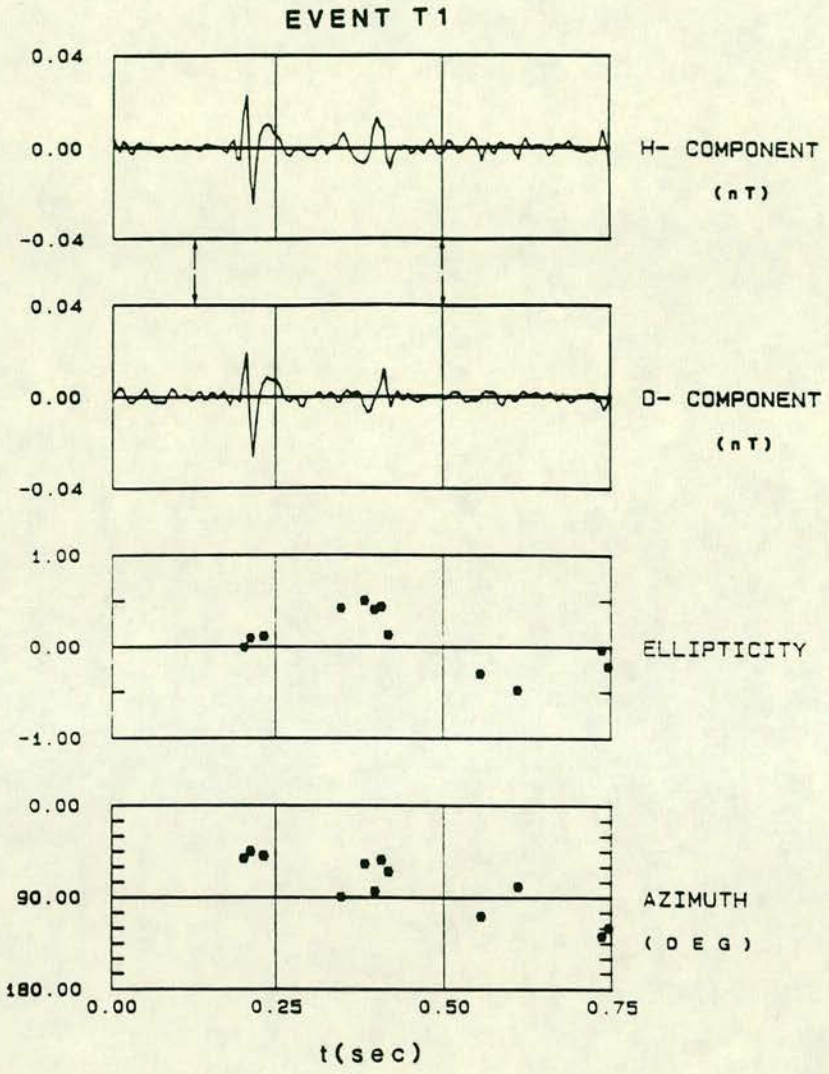


Fig. 4. Three examples of the time domain polarization analysis of impulsive transients. *H* and *D* correspond to the N–S and E–W magnetic components; real time separation of events T1–T2, 90.75 s, events T2–T3, 63.75 s. (*Continued over.*)

spectra of the impulsive sferics, annotated T1, T2 and T3. The spectra have been evaluated from the complex signal using the Maximum Entropy (ME) method (BURG, 1975). The high resolution capability of ME spectra for short data lengths permits the analysis of individual waveforms without interference from other sources. Thus only segments of the available data have been analysed. The segments are indicated in Figs. 3 and 4 with thin arrows.

It is immediately apparent that the spectral signatures of the three background events of Fig. 3 are very different. Recall that the example waveforms all possess different azimuths; in particular, events B1 and B3 are almost orthogonal. Our results indicate a variation of the spectral signature with azimuth. The same observation can be made for the impulsive transient events T1, T2 and T3. As indicated previously, the shape and trend of the spectral curves can be used to estimate the source–receiver separation. Care must be exercised while making such interpretations using the background field, as the theoretical results of JONES and KEMP (1970, 1971) refer to the excitation of the cavity by extra-large single discharges. In addition, the work assumes relatively simple source spectra. The spectra of the background and transient events (more representative of the above

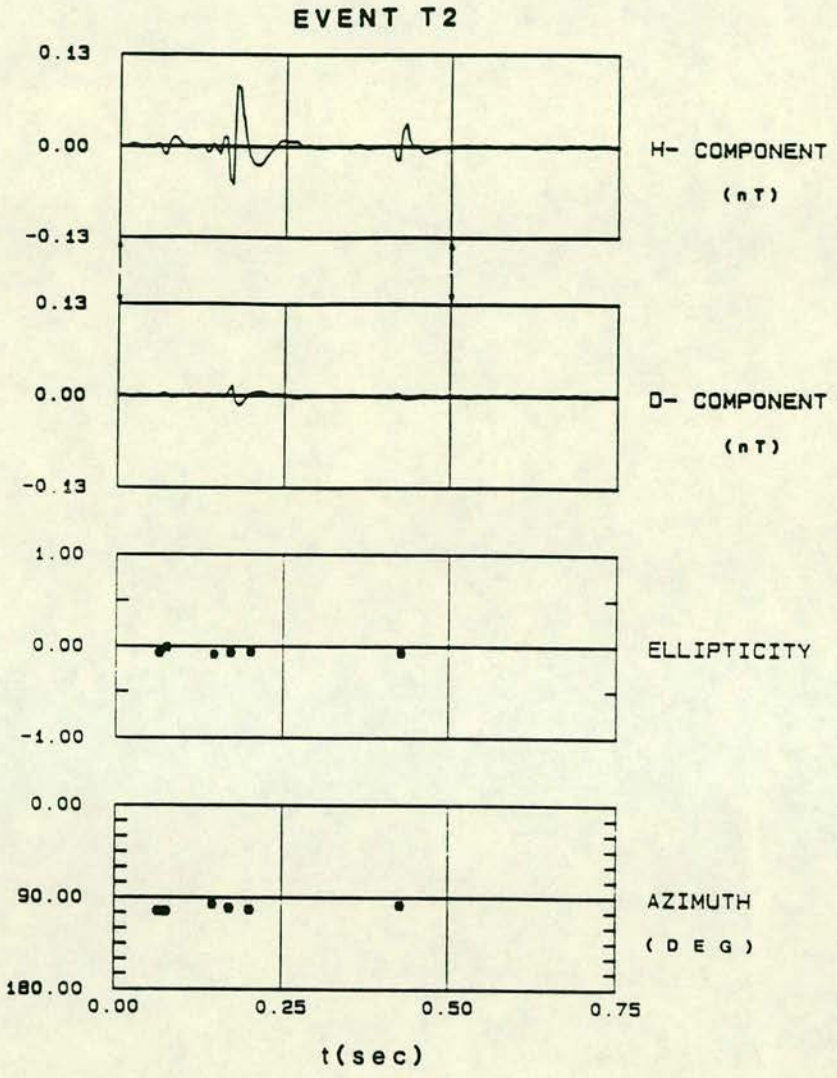


Fig. 4 continued. (*Continued over.*)

authors' models) display qualitative differences. The former possess sharper peaks and larger peak-to-trough ratios, possibly, because of the more severe attenuation experienced by weak sources in the cavity. However, events B2 and B3 display simple waveforms and do not appear to be the result of multiple sources. It must also be noted that our instrumentation, primarily intended for crustal sounding, includes a notch filter that was centred on 47.5 Hz. The notch was operational when the background events were collected. Frequencies above 40 Hz are therefore attenuated for spectra B1, B2 and B3.

With every caution we assert that the spectrum of event B2 is very similar in shape to the theoretical results of JONES and KEMP (1970, fig. 7) for a range (D) of 10,500 km. Event B3 possesses a spectrum which indicates a source less than 8000 km away. With regard to the impulsive transients, it can be inferred with confidence that event T2 represents a distant soruce ($D > 10{,}000$ km), whereas event T3 may represent a range (D) of less than 7000 km. The results presented above have been found to be repeatable. Waveforms selected on the basis of common azimuths display similar shapes and spectral signatures. We observe an association of spectral variability with azimuth and widely separated sources. We interpret

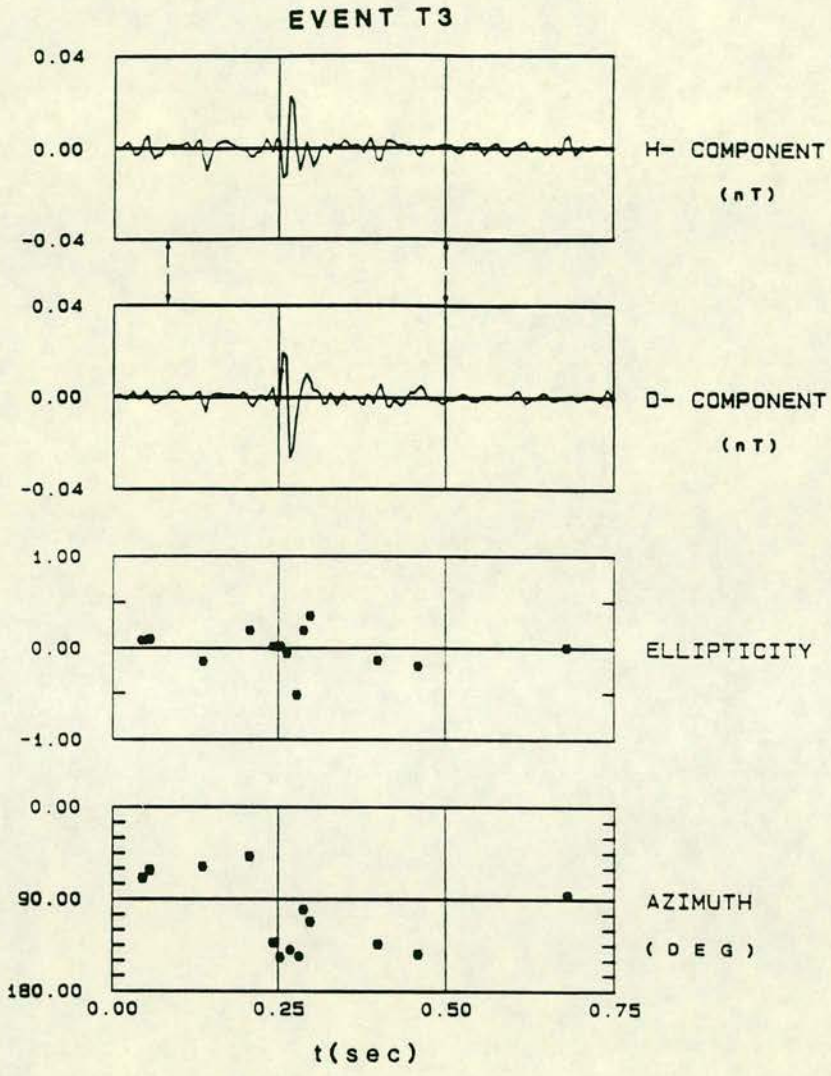


Fig. 4 continued.

this as being due to signals being received from different source locations and travelling over different paths.

It should be noted that the data used in the present study are not ideally suited to the determination of source locations. Such determinations are better made using spectra of the wave impedance, requiring measurements of the vertical electric field (KEMP, 1971; BLIOKH *et al.*, 1980). Our main conclusions arise from the results obtained from our time domain study of the polarization properties of the waveforms. The results show that many of the waveforms exhibit a stable linear polarization over several cycles. This behaviour can be interpreted as the direct arrival of a short path signal from a given source location. Such properties are most clearly exhibited by the larger transient events, but are also observed at the background level. We conclude that thunderstorm activity giving rise to sources which excite the Earth–ionosphere cavity can be simultaneous and therefore incoherent, but the effects from single sources are detectable even at the background level.

Acknowledgement—This paper is published with the approval of the Director, British Geological Survey (NERC).

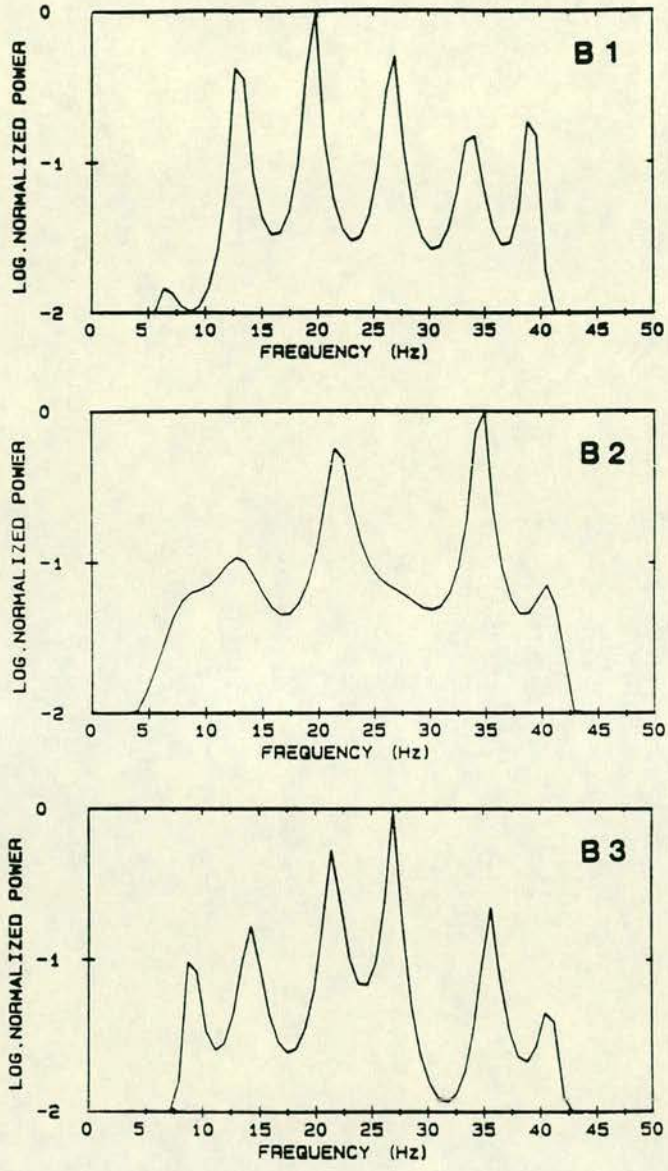


Fig. 5. The spectral signatures of the example sferics. Maximum Entropy power spectra of the azimuthal magnetic field. Data lengths analysed: event B1, 52 points (0.26 s); event B2, 48 points (0.24 s); event B3, 52 points (0.26 s) [filter length $M = 20$ used throughout]; event T1, 76 points (0.38 s); event T2, 100 points (0.50 s); event T3, 85 points (0.425 s) [filter length $M = 30$ used throughout]. (*Continued over.*)

REFERENCES

BEAMISH D. and TZANIS A.	1986	*J. atmos. terr. Phys.* **48**, 187.
BLIOKH H., NIKOLAENKO A. P. and FILIPPOV YU. F.	1980	*Schumann Resonances in the Earth–Ionosphere Cavity* (LLANWYN-JONES D., Ed.), IEE Electromagnetic Wave Series 9. P. Peregrinus Ltd, Stevenage.
BORN M. and WOLF E.	1959	*Principles of Optics*. Pergamon Press, New York.
BURG J. P.	1975	Ph.D. Thesis, Stanford University, Stanford, California.
FOWLER R. A., KOTICK B. J. and ELLIOTT R. D.	1967	*J. geophys. Res.* **72**, 2872.
GABOR D.	1946	*J. Instn elect. Engrs* **93**, 429.

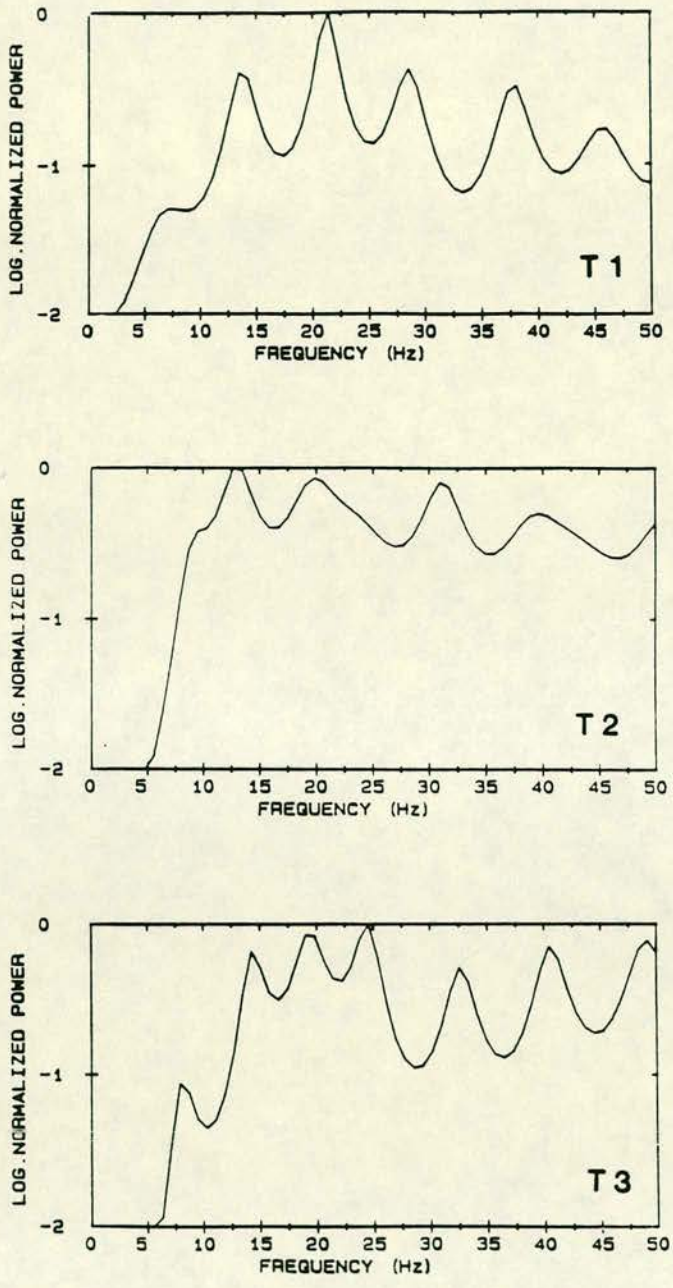


Fig. 5 continued.

Jones D. L. and Kemp D. T.	1970	*J. atmos. terr. Phys.* **32**, 1095.
Jones D. L. and Kemp D. T.	1971	*J. atmos. terr. Phys.* **33**, 557.
Kemp D. T.	1971	*J. atmos. terr. Phys.* **33**, 919.
Kodera K., Gendrin R. and De Villedary C.	1977	*J. geophys. Res.* **82**, 1245.
Madden T. and Thomson W.	1965	*Rev. Geophys.* **3**, 211.
Turman B. N. and Edgar B. C.	1982	*J. geophys. Res.* **87**, 1191.
Tzanis A. and Beamish D.	1986	Geomagnetism Research Group, Internal Report 86/13, British Geological Survey.

J Geophys (1987) 61:97–109

Journal of Geophysics

Audiomagnetotelluric sounding using the Schumann resonances

A. Tzanis* and D. Beamish

British Geological Survey, Murchison House, West Mains Road, Edinburgh EH9 3LA, United Kingdom

Abstract. The Schumann resonance waveforms in the lower ELF band (5–100 Hz) are produced within the Earth-ionosphere cavity by distant lightning discharges; they provide a useful source field for shallow audiomagnetotelluric (AMT) crustal sounding. In this study we investigate their waveform characteristics that are important to the assumptions of AMT sounding. A time-domain polarization analysis technique is applied to a variety of examples of Schumann resonance waveforms. The multiplicity of worldwide thunderstorm centres provides a background activity which is generally incoherent and, accordingly, displays incoherent polarization characteristics. Superimposed on the background are larger-amplitude transient events (sferics) from individual thunderstorm centres; they represent the response of the Earth-ionosphere cavity to very large lightning discharges and are generally linearly polarized at a given azimuth. The analysis indicates that the Schuman resonance waveforms provide a plane-wave source field, as required for electromagnetic crustal sounding. The differences in the "received" characteristics between the two waveform types prompted an investigation of the extent to which the waveform type and its particular polarization characteristics influence the determination of a geoelectric sounding curve. A detailed study, carried out with a multivariate maximum entropy spectral analysis algorithm, indicates that the two types of Schumann resonance waveform provide repeatable and consistent results at the 95% confidence level and that the linear polarizations associated with the sferics do not influence the estimation of the Earth response.

Key words: Electromagnetic induction – Audiomagnetotelluric – Schumann resonances – Polarization analysis – Autoregressive spectral estimation

Introduction

Audiomagnetotelluric (AMT) crustal sounding is an extension of the lower frequency ($f<1$ Hz) magnetotelluric (MT) method to higher frequencies. The basic requirement for both methods of geoelectric sounding is a time-varying electromagnetic source field. Such an external field provides the energy source for electromagnetic induction in the conducting Earth. The time variations are recorded as two components of the induced electric field (E_x, E_y) and the combined inducing and induced magnetic field (H_x, H_y) at a particular location. Linear relationships between the field components, calculated as a function of frequency, enable a geoelectric sounding curve to be obtained and interpreted. High-frequency source fields enable relatively shallow crustal targets to be explored. If the instrumental bandwidth extends to 100 Hz, skin depths would range from 158 m to 16 km in uniform crustal rocks with resistivities from 10 to 10^5 ohm·m, respectively.

The transition frequency from MT to AMT sounding is not arbitrary (Keller, 1971). A frequency of several hertz is associated with a distinct minimum in the Earth's natural electromagnetic source spectrum (Watt, 1967). At frequencies of $f<1$ Hz the source fields are primarily of magnetospheric and ionospheric origin, i.e. associated with perturbations of the main geomagnetic field. The geomagnetic energy spectrum for $f<1$ Hz, observed in active regions of the magnetosphere, displays an approximately f^{-2} to f^{-3} dependence (Russell et al., 1970). This spectral decay, combined with the low energy transmission coefficient of the ionosphere at such frequencies (Abbas, 1968), suggests that only a few rare magnetospheric disturbance fields will contribute significant amounts of energy at frequencies greater than a few hertz. The natural noise spectrum in the 5 Hz–25 kHz range is primarily due to electromagnetic perturbations generated by lightning discharges (ELF sferics). Such sources may contain a wide spectrum of frequencies. However, for distant (i.e. worldwide) sources, the energy undergoes multiple reflection in the Earth-ionosphere waveguide giving rise to frequency-dependent attenuation. In the lower audio band (i.e. $f<2$ kHz) the energy propagates as a transverse magnetic mode in which the electric field is largely radial. The largest energy components in this lower audio band are the Schumann resonance modes. The first five of these modes have typical frequencies of 7.8, 14.1, 20.3, 26.3 and 32.5 Hz (Madden and Thomson, 1965). The present study considers some aspects of sferic source fields, in particular the Schumann resonance modes, and their utilisation in AMT crustal sounding.

Although AMT instrumentation may use the natural spectrum up to 10 kHz (Strangway et al., 1973), we confine our attention to the low-order Schumann resonance modes ($f<100$ Hz). Telluric and magnetotelluric prospecting methods, using the fundamental mode ($f=7.8$ Hz), have been described by Slankis et al. (1972). The analysis of the characteristics of audio signals described by Telford (1977) was undertaken using time-averaged measurements of the

Offprint requests to: D. Beamish

induced electric field. This would appear to be the least effective way to determine the source characteristics of signals propagating in a transverse magnetic mode. Although AMT tensor measurements are now increasingly common, descriptions of source (and noise) characteristics are usually restricted to the spectral content of the waveforms (Hoover et al., 1978; Labson et al., 1985). In order to provide a more detailed description of typical source characteristics, we investigate here a useful method for the polarization analysis of sferic waveforms.

Common assumptions in both the MT and AMT methods are that the source field approximates an incident plane wave and that no current sources exist within the region through which electromagnetic induction takes place. It is quite possible for both of these assumptions to break down when we attempt to use sferic sources in the lower audio band (5 Hz–2 kHz). The plane-wave assumption is not appropriate for energy sources which derive from local and regional thunderstorm activity (examples are provided in Keller and Frischknecht, 1966). The power distribution grid and its users are capable of providing both narrow-band and broad-band current sources which invalidate the second assumption. The present study avoids the second complication by using "clean" Schumann resonance waveforms (Beamish and Tzanis, 1986). The first assumption is examined by considering the general characteristics of sferic sources across the Schumann bandwidth. A polarization analysis of the magnetic field components of Schumann resonance waveforms is then described and their ability to provide repeatable geoelectric sounding curves is then investigated.

General characteristics of sferic sources

In the ELF band, electromagnetic energy is largely produced by radiation from vertical lightning discharges. Descriptions of the processes involved can be found in Uman (1969). For a particular model of the discharge (e.g. single or return strokes), the discharge will be characterized by its current moment (time domain) or equivalently by its source spectrum (frequency domain). The electric field energy generated comprises electrostatic, inductive and radiation components which decay with distance (D) as $1/D^3$, $1/D^2$ and $1/D$, respectively. The electrostatic term represents the dominant component up to distances of about 20 km, while the radiation term predominates at distances in excess of 100 km (Bliokh et al., 1980). In order to observe the Schumann resonances, we require that the source-receiver separation be large (e.g. >1000 km) and that the source provides forced oscillations within the Earth-ionosphere cavity. Under these circumstances, the received spectrum will be the product of the source spectrum and the waveguide transmission function.

The three main equatorial thunderstorm centres in Central/South America, Central Africa and S.E. Asia (Keller and Frischknecht, 1966) provide a nearly uniform source of ELF noise in that there exists a high probability that a thunderstorm is in progress somewhere on the globe. The presence of such recurring centres of activity suggests that mid- and high-latitude locations maintain a high probability of being in the far-field ($D>10000$ km) region with respect to the main source locations. In addition to this spatial distribution, thunderstorm activity can also be characterized by a time dependence. The broad feature of late afternoon maximum (20:00 U.T.) and a night-time minimum (04:00 U.T.) in ELF noise may be further developed by including the time-dependent contributions from the three main thunderstorm centres. The activity of these centres, as a function of universal time, is reproduced from Bliokh et al. (1980) in Fig. 1. In this *statistical* model, ELF energy would be received from each of the three centres during the course of the day.

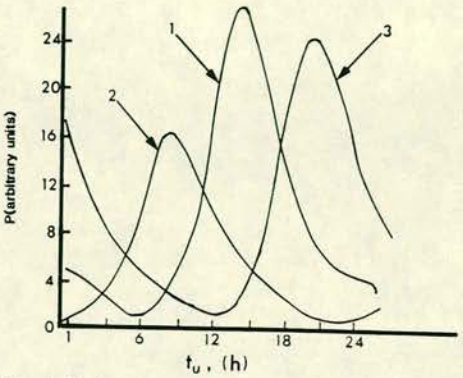


Fig. 1. Lightning activity of the global thunderstorm centres as a function of universal time. (1) African centre, (2) Asiatic and (3) American. After Bliokh et al. (1980)

According to Bliokh et al. (1980), it is common to distinguish two main components of the noise in the lower audio band. The first component is the continuous background level due to the global sum of all (incoherent) sources. The spectral variance of such noise is remarkably stable and remains almost constant over time intervals of the order of 1–2 h; it gives rise to the diurnal pattern noted above. The second main component is an ELF transient event which is the response of the cavity to a single, very powerful lightning discharge. The amplitude of such events can exceed the background level by a substantial factor (e.g. >10) and their duration is typically a few cycles. Such events may be coherent over most, if not all, of the Earth. It is these two main components that we will consider in the present study.

The structure of the ELF spectrum (i.e. the relative amplitude and the position of spectral peaks) is strongly dependent on the range from source to receiver. With simplifying assumptions concerning the source spectrum, the range of a lightning discharge is determined by comparing the recorded spectrum with spectra calculated for different ranges (Jones and Kemp, 1970). To locate a lightning discharge it is then sufficient to obtain the angle of arrival (or bearing) of the wave from the two horizontal magnetic components (Kemp, 1971). If the source spectrum is not simple (i.e. it may comprise multiple return discharges), the received spectrum will be highly structured and the above procedure is difficult to apply. This particular problem can be overcome by using the spectrum of the wave impedance (E_r/H_θ), if all three electromagnetic components are measured (Kemp and Jones, 1971; Ingmann et al., 1985). We note that if measurements of the radial electric field (E_r) were routinely included in the AMT instrumentation, the exclusion of local or regional sources would, in theory at least, be possible. If E_r is not observed, the simplest source characteristic to investigate is the bearing of the wave by a

polarization analysis of the two horizontal magnetic components.

Polarization analysis of Schumann waveforms

Some of the data used in the present analysis were used in the spectral study of Beamish and Tzanis (1986). The instrumental conditioning and collection scheme provides four decades of data from 100 to 0.01 Hz. Decade 1 is sampled at 400 Hz and resampled at 200 Hz, providing a data bandwidth from 100 to 10 Hz. This decade attenuates the fundamental (7.5 Hz) Schumann resonance mode but provides data for studies of the higher-order modes. It is data from this decade that are used in the present study. The data were collected in western Anatolia (Turkey), geographic coordinates 40.5° N, 30° E. The spectral content of typical decade-1 data has been described by Beamish and Tzanis (1986).

A large number of individual data windows were collected at a number of sites. We present detailed examples of waveform characteristics on two separate days, at two separate locations. The three examples from the first day (8 June 1984, between 10:40 and 11:00 L.T.) have relatively uniform variance and can be considered representative of the ELF background field. They are referred to as events B1, B2 and B3. The three examples from the second day (27 July 1984, between 10:37 and 10:50 L.T.) are impulsive waveforms having peak-to-peak amplitudes of between $\times 5$ and $\times 15$ that of the background level and are considered representative of typical ELF events. They are referred to as events I1, I2 and I3. The waveforms recorded can display a high degree of amplitude/frequency variance as typified in the four equivalent sonograms of Fig. 2. This figure shows the spectral content of successive data windows in the N–S horizontal magnetic field of decade-1 data. In general, the Schumann resonance waveforms comprise "events" superimposed on a background. In order to extract realistic polarization parameters for the waveforms considered, it has been necessary to investigate time-local polarization analysis techniques.

In the present study we adopt a technique, proposed by Kodera et al. (1977) and first described by Gabor (1946), which permits the derivation of the polarization parameters in the time domain in the form of instantaneous values. This procedure has the advantage of tracing any variation in the state of polarization of *individual* waveforms encountered within a particular data window. The technique extends the formalism obeyed by a complex signal $c = x + iy$, associated with any two orthogonal components x and y, to a two-component representation. Kodera et al. (1977)

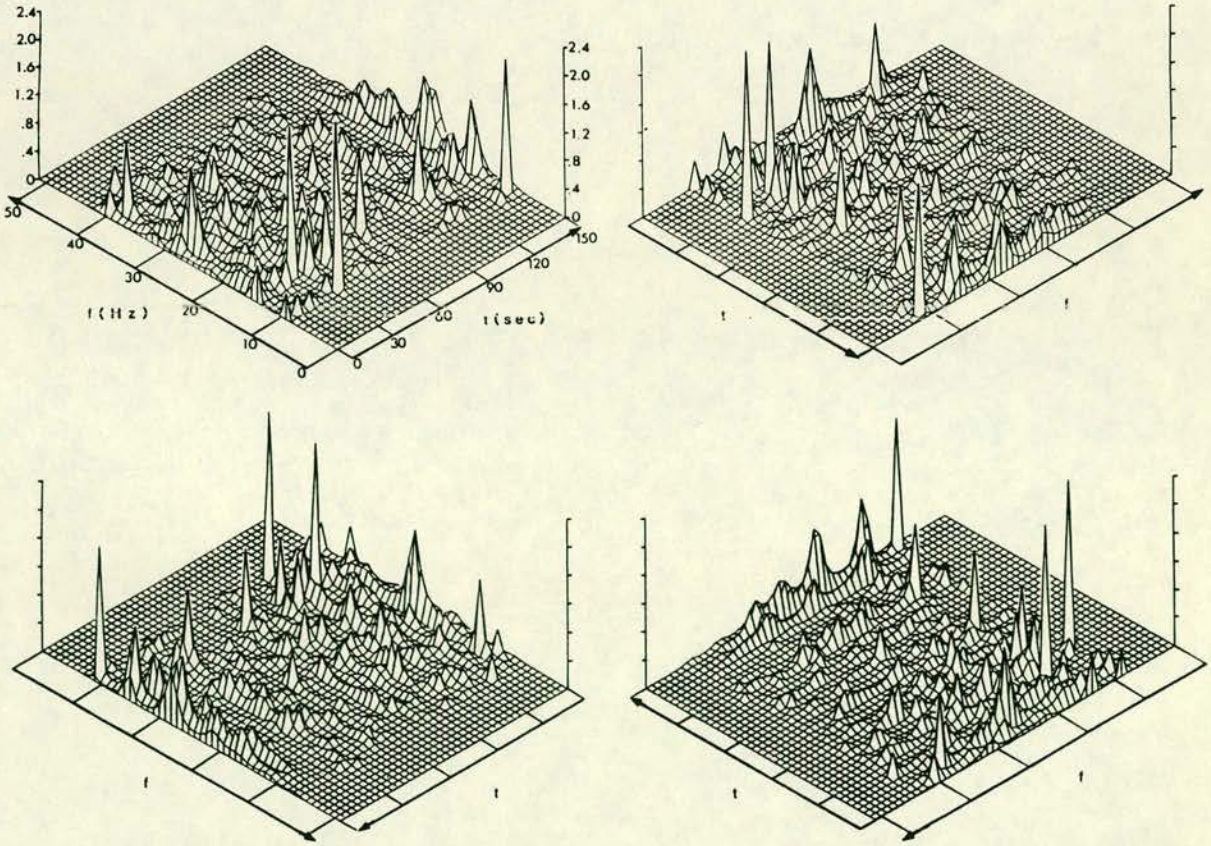


Fig. 2. High-resolution sonogram of the N–S (H_x) magnetic component. Four perspective view of the *same* sonogram. Data collected on 8 June 1984. Power density values in $(nT)^2\ Hz^{-1} \times 10^{-6}$. Decade-1 data, 10–100 Hz. Maximum entropy power spectra. Filter length $M = 35$

demonstrate that it is possible to decompose the complex signal in terms of circular components rotating in opposite directions. At any given frequency f_0, we may write:

$$c(t, f_0) = C(w_0) \exp(i w_0 t),$$

which represents a circular harmonic in the complex plane, rotating in a sense which depends on the sign of w_0. Positive (+) frequencies correspond to circular components rotating in the counterclockwise direction and negative (−) frequencies to components rotating in the clockwise sense. The power spectrum of the complex signal is related to the intensity of the circular components and the spectral asymmetry to the polarization of the real signals. The authors detail how the two complex circular decomposites of the complex signal can be written in terms of the equivalent real decomposites of the two orthogonal components, in the form:

$$c^+(t) = x^+(t) + iy^+(t),$$
$$c^-(t) = x^-(t) + iy^-(t).$$

By introducing the concept of the complex *analytic* signal associated with any real signal (e.g. Ville, 1946), one can easily show that:

$$x^+(t) = x_a(t)/2 \qquad x^-(t) = x_a^*(t)/2$$

and similarly for the y component. The subscript a denotes the analytic signal; the asterisk, complex conjugation. Consequently, one can deduce that:

$$c^+(t) = [x_a(t) + iy_a(t)]/2,$$
$$c^-(t) = [x_a^*(t) + iy_a^*(t)]/2.$$

For monochromatic or quasi-monochromatic signals, the standard polarization parameters are easily obtained as:

half major axis: $a = |c^+| + |c^-|$

half minor axis: $b = |c^+| - |c^-|$

ellipticity: $\varepsilon = b/a$

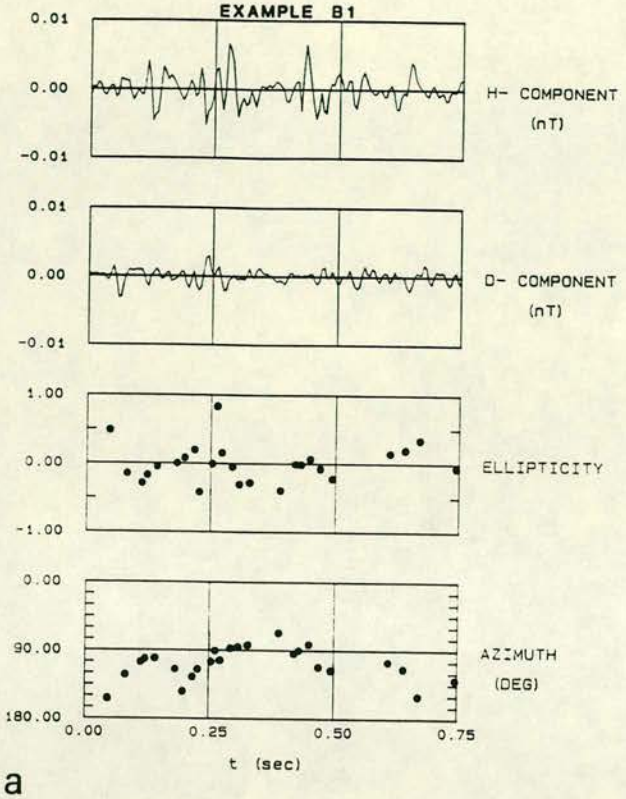
azimuthal angle: $\theta = [\arg(c^+) + \arg(c^-)]/2.$

In the above expressions, b and ε are positive for counterclockwise rotation and negative for clockwise rotation. We use a different numerical implementation to that suggested by Kodera et al. (1977), which appears to be more suitable for data displaying large waveform variability. This has been described in Tzanis and Beamish (1986b).

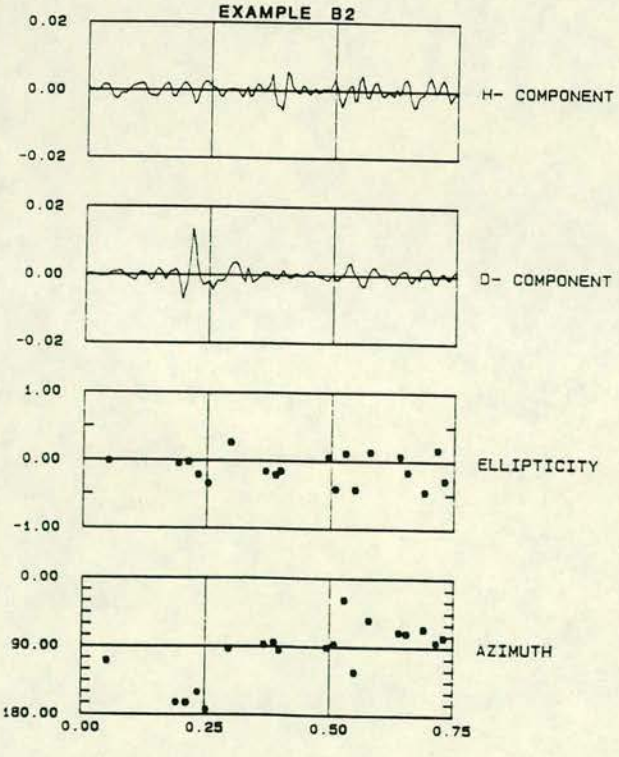
The same method of presentation is repeated for all six examples. Events B1, B2 and B3 (background field) consist of 150 data points in (H_x, H_y). The time interval considered is 0.75 s. Events I1, I2 and I3 consist of 50 data points (0.25 s) in (H_x, H_y). For each example, the two-component data are displayed first. The polarization parameters ellipticity (ε) and azimuth (θ) are then displayed in the time domain, below the corresponding data. The instantaneous values of ε and θ consist of a series of discrete symbols that correspond to the extrema in (H_x, H_y), i.e. the signal inflexion points, evaluated using the modulus. If the amplitude of the modulus falls below a pre-set threshold (usually taken as a percentage amplitude of the maximum excursion encountered), the result is omitted. This is a simple and effective method of rejecting noise. The thresholds used here are of the order of 25% for the background field and 10% for the ELF transient events. In keeping with the previous definitions, a value of $\varepsilon = 1$ corresponds to a circular wave rotating in a counterclockwise (+) sense, a value of $\varepsilon = 0$ corresponds to a linearly polarized wave and a value of $\varepsilon = 1$ to a circular wave rotating in a clockwise (−) sense. Azimuths are defined as magnetic west (180°), magnetic north (90°) and magnetic east (0°).

Figure 3a shows the results from the analysis of event B1. There are no strongly dominant waveforms and the peak-to-peak amplitudes are less than 0.01 nT. A high degree of variability in polarization characteristics is apparent. Figure 3b shows the results obtained from event B2. Peak-to-peak amplitudes are all less than 0.02 nT. The large H_y-component waveform is clearly linearly polarized, while the other waveforms display a variety of polarization characteristics. Example B3 (Fig. 3c) is not totally representative of the background field due to the presence of a large-amplitude waveform (just after 0.50 s), possessing a larger variance level and more stable polarization characteristics as a function of time. This example can be directly compared with the impusive waveforms shown in Fig. 4. It is easily recognized that the three main impulsive waveforms shown in Fig. 4 all posses a near-zero initial ellipticity but distinctly different azimuths. The azimuths of examples I1 and I2 differ by 66°; the azimuths of I2 and I3 by 60°. All waveforms possess similar amplitudes. The real time interval between I1 and I3 is approximately 8 min.

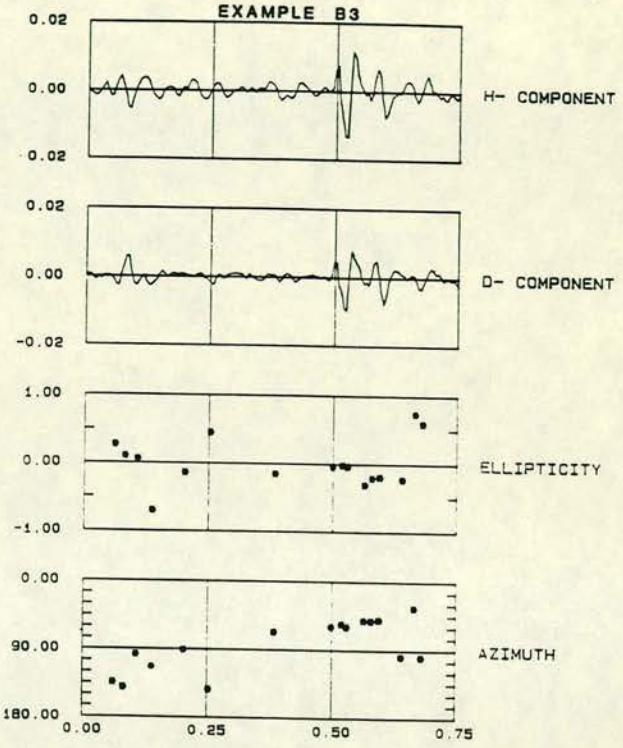
Our purpose in presenting the above results has been to provide typical examples of the polarization characteristics of ELF waveforms. The spectral content of these data indicate that the waveforms are due to the Schumann resonances, originating in the Earth-ionosphere cavity (Beamish and Tzanis, 1986). The most straightforward use of the polarization characteristics presented is the identification of the bearing of the source. The horizontal bearing of the source is perpendicular to the polarization azimuth of the horizontal magnetic field. The bearings of the background waveforms display the prescribed level of variability for incoherent sources with two exceptions. The two larger-amplitude waveforms in Fig. 3b and c are both characterised as short-duration, quasi-linear waveforms. Their amplitude excursion exceeds the background by a factor of less than 5 and the bearings of the two sources appear in opposite quadrants. Their polarization characteristics appear to be a more stable function of time. The three main impulsive events considered in Fig. 4 appear a simple, damped quasi-sinusoidal oscillations. These waveforms posses the characteristic commencement of typical transient events, interpreted by Jones and Kemp (1970) as the response of the cavity to single, powerful and relatively short-duration lightning discharges. The initial portion of any ELF disturbance is due to the direct reception of radiation from the discharge. Subsequent fluctuations are due the circulation of energy around the globe. For genuinely impulsive waveforms, the onset should be characterised by an ellipticity of zero. This feature is common to the three transient events analysed. The two "main" waveforms encountered in Fig. 4b are a good example of source "repetition" on a time scale of 0.25 s. The same effect is also evident in Fig. 4a. Another interesting feature in this figure is the existence of a small-amplitude waveform just prior to the onset of the main impulsive event, with distinctly different polarization characteristics ($\theta = 87°$, $\varepsilon = 0.25$). The results demonstrate the extremely time-local behaviour of the ELF sferic waveforms. When the bearings of the three impulsive forms of Fig. 4 are compared, it is apparent that the sources have different azimuthal locations. The study by Tzanis



a



b



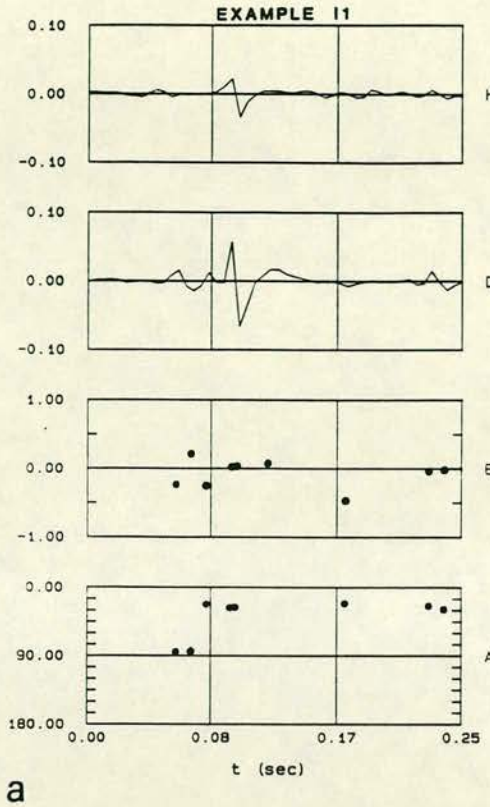
c

Fig. 3a–c. Three examples of the time-domain polarization analysis of background sferics. *H* and *D* correspond to the N–S and E–W magnetic components. **a** example B1; **b** example B2; **c** example B3

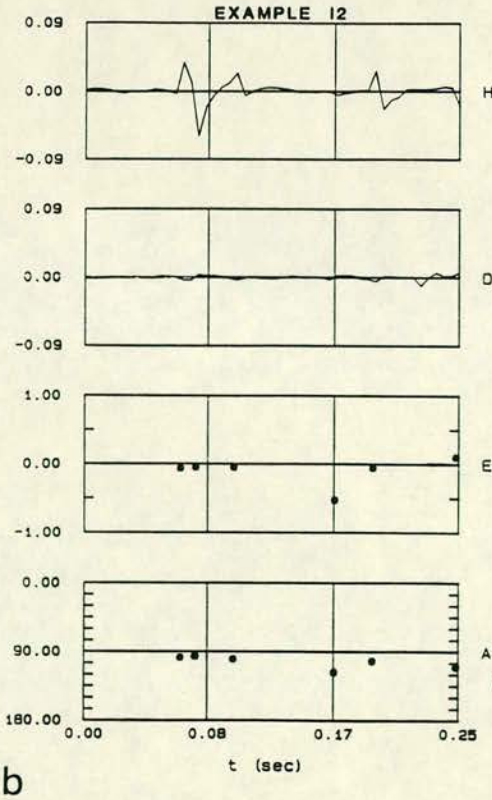
and Beamish (1986a) indicates that the source locations of such waveforms possess ranges (*D*) of the order of several thousand kilometres and are widely distributed across the globe. The same is evident from the data presented in this analysis; individual sources are detectable even at the background level (e.g. examples B1 and B2). The low *Q* factors (less than 8) calculated from the power spectra of the ELF sferics also indicate that energy has propagated over large distances in the Earth-ionosphere cavity. They reflect the effect of the low transmission coefficient of the cavity resonator (Bliokh et al. 1980) and can be used as a diagnostic aid in verifying the nature of the received waveforms. In conclusion, the above observations point towards the validity of the plane-wave assumption when using 'typical' Schumann resonance waveforms.

Geoelectric sounding curves from single events

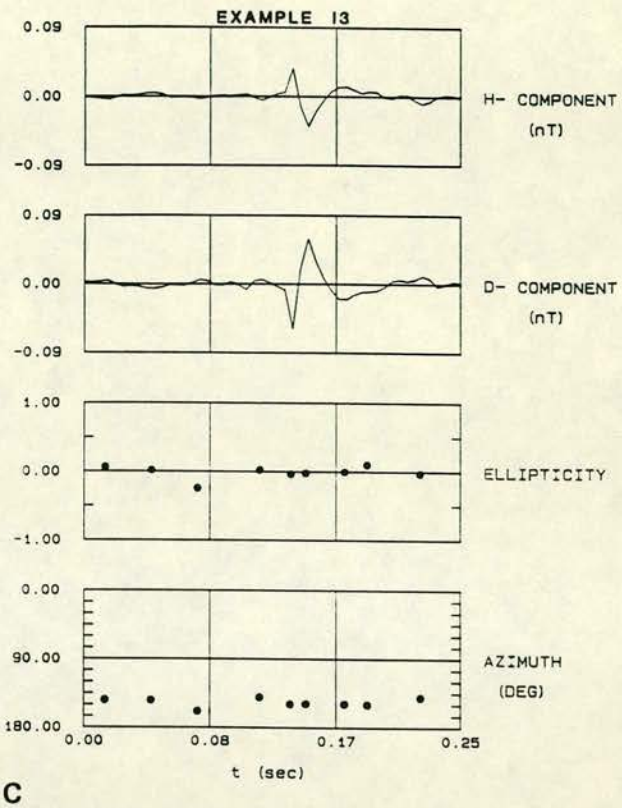
In general terms it appears that the Schumann resonances provide two main types of source waveforms for use in geoelectric sounding. The background source fields possess uniform, but low, variance and provide a complete range of source azimuths and ellipticities on time scales equivalent to the duration of the waveforms. The large-amplitude transient events are of short duration (typically less than 0.25 s) and appear as quasi-linearly polarized waveforms at a given



a



b



c

Fig. 4a–c. Three examples of the time-domain polarization analysis of ELF transient events. *H* and *D* correspond to the N–S and E–W magnetic components. **a** example I1; **b** example I2; **c** example I3

azimuth. The transient "events" available in the data provide a wide variety of such azimuths. Impulsive sources are clearly valuable for geoelectric sounding in that they possess a substantial increase in the signal-to-noise (S/N) ratio when compared with background sources. At least two source-field polarizations are required to adequately sound general 3-dimensional geoelectric structures. The use and effects of polarized source fields has been described by Hermance (1973), Lienert (1980) and Pedersen and Svennekjaer (1984). From the preceding discussion it is evident that the Schumann resonance waveform characteristics are associated with distinct azimuthal locations and propagation paths. We now proceed to investigate the extent to which such waveforms can provide repeatable geoelectric sounding curves, i.e. whether the waveform type and its particular polarization characteristics can possibly influence the determination of the Earth response. We are *not* concerned here with noise effects; thus, we deliberately choose examples with small noise content. A high (frequency-time) resolution technique is applied to sferic waveforms. This enables the response due to *individual* waveforms, directly related to a given discharge source and propagation path, to be determined.

Data analysis and spectral estimation procedures

The data analysis follows the conventional procedures described by Vozoff (1972). The electric and magnetic field

components are related, in the frequency domain, by the two equations:

$$E_x(w)=Z_{xx}(w)\,H_x(w)+Z_{xy}(w)\,H_y(w), \quad (1)$$

$$E_y(w)=Z_{yx}(w)\,H_x(w)+Z_{yy}(w)\,H_y(w) \quad (2)$$

or, in matrix form, as $E=Z\cdot H$, where $E=[E_x\ E_y]^T$, $H=[H_x\ H_y]^T$ and Z is termed the impedance tensor. The two pairs of impedance elements are usually estimated by least-squares solutions which minimize noise on a particular data channel (Sims et al., 1971). The quality of the least-squares solutions for each of the the two pairs of impedance elements is then obtained from the multiple (predicted) coherence function between a measured electric field component and that predicted by the least-squares solution.

The auto- and cross-spectral estimates between the field components are conventionally computed using the Fast Fourier Transformation (FFT) technique. Given the inherent variance of the resulting raw spectral density function, smoothing over a frequency interval (often arbitrarily chosen) is usually prescribed in order to stabilize the spectral density estimates. The frequency resolution of the resulting impedance estimates is thus drastically reduced. Additional disadvantages of the procedure are that the resulting (smoothed) spectral estimates contain the contributions of several (and often unconforming) frequency-local properties over the interval of smoothing, and that the FFT cannot handle short data lengths without loss of statistical significance in the spectral estimator. To obtain highly resolved and statistically robust impedance estimates that display properties localized in frequency, a different spectral approach can be adopted.

It appears that for the case of spectral estimation from a stationary time series, there exists a ubiquitous possibility that we can represent the time series by an autoregressive (AR) process of the form:

$$x(t)=\sum_m a(m)\,x(t-m)+e(t)$$

(Ulrych and Bishop, 1975; Jaynes, 1982), where $e(t)$ is a white-noise error series and $a(m)$ is an absolutely summable filter. Using such a model, it is possible to improve frequency resolution by determining the spectrum from the properties of the filter $a(m)$, $m=0,\ldots,M$ that best adapts to the given data set. The problem of spectral estimation then reduces to that of determining the optimum filter coefficient vector. Burg (1968) proposed a method for estimating the filter coefficients and hence the power spectrum of a stationary time series in the maximum entropy (ME) sense. Beamish and Tzanis (1986) demonstrated the capacity of the algorithm for the production of high-resolution spectra when applied to single-component Schumann resonance waveforms (i.e. the univariate case). In order to apply the same spectral technique to impedance estimation we require to calculate both auto- and cross-spectral estimates (i.e. the multivariate case).

Consider a vector time series of the form:

$$x(t)=[x_1(t)\,x_2(t)\,\ldots x_p(t)]^T, \quad t=1,\ldots,N$$

consisting of p simultaneous data channels. The equivalent linear AR system will now assume the form:

$$x(t)=\sum_m a^T(m)\,x(t-m)+e(t)$$

where $e(t)$ is a p vector white-noise series and a is a $p\times p$ vector absolutely summable filter. The power density spectrum will be given by the expression:

$$s(z)=\Delta t\cdot[A(z)^{-1}]^{*T}\cdot P_m\cdot[A(z)^{-1}]$$

where P_m is the $p\times p$ vector residual error power and $A(z)$ is the $p\times p$ vector z-transform of the filter a. Δt is the data sampling rate and the asterisk denotes complex conjugation. The least-squares minimization of $e(t)$, to provide the optimum unit prediction error filter a, has been considered by a number of authors (e.g. Strand, 1977; Morf et al., 1978) in a more or less direct generalization of Burg's algorithm. The simultaneous treatment of p data channels provides the opportunity for a direct evaluation of both auto- and cross-spectral components using the $p\times p$ operator a.

For our data consider the vector time series

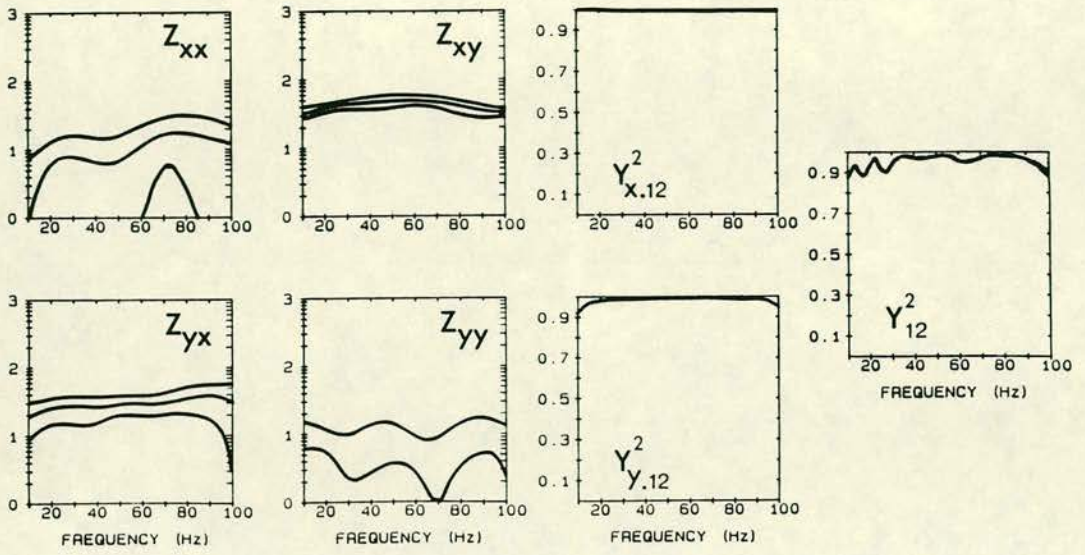
$$x(t)=[E_x(t)\ E_y(t)\ H_x(t)\ H_y(t)]^T, \quad t=1,\ldots,N,$$

consisting of four data channels. The above spectral procedure is then used to form the spectral components required for the least-squares solutions. The spectral analysis algorithm implemented is a modification of the one due to Strand (1977). The estimation of the error associated with the resulting impedance tensor elements was performed according to Pedersen (1982). In the latter, the statistical significance of the principal-component spectral density estimates involved is introduced through their associated number of degrees of freedom; their distribution is assumed to be complex normal for the raw spectral density matrix and complex Wishart for its smoothed equivalent (e.g. Priestley, 1981, pp. 693–701). This is true for spectra calculated with the conventional techniques; however, the multivariate ME spectral estimator lacks an exact statistical description of its properties.

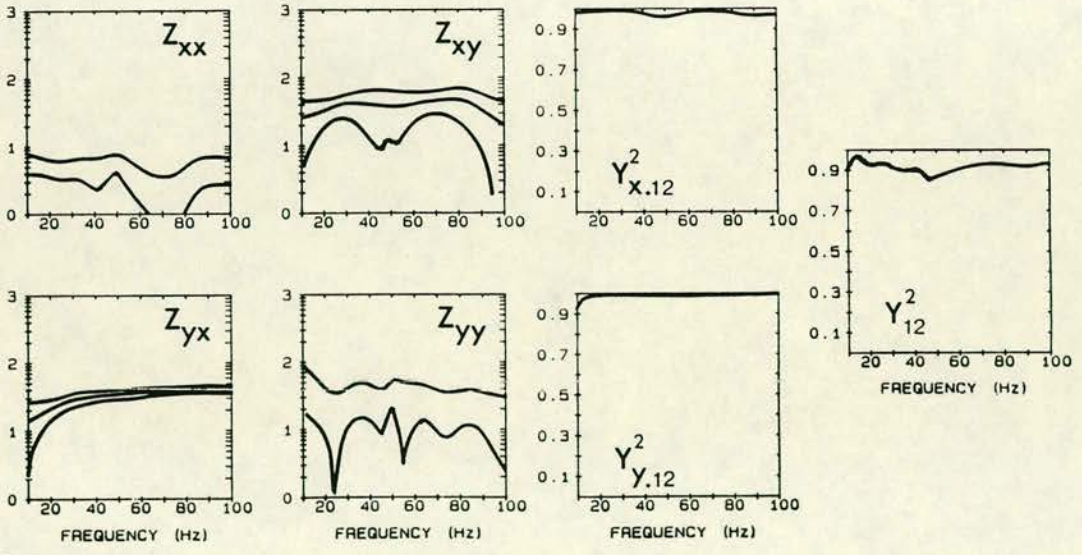
Progress can be made if we consider the statistics of the generalized linear regression system. Its spectral density matrix is shown to comprise a class of consistent, asymptotically unbiased and asymptotically complex normally distributed estimates (e.g. Brillinger, 1981, Chapter 8). It is therefore conceivable that the multivariate AR spectral estimator, being a particular case of such a system, will possess similar statistical properties, although the moments of the distribution are yet to the specified. Such an argument is based on, and enhanced by, the fact that all the entries in the data vector x are assumed to be second-order stationary time series, jointly normally distributed; this somehow prescribes the result. The number of degrees of freedom associated with the principal components of the spectral density matrix are taken to be $n=N/M$, as a direct generalization of the result by Kromer (1970) concerning the statistics of the univariate AR spectral estimator. The latter was found to be unbiased, consistent and asymptotically normally distributed. The above arguments provide, at best, an approximation. However, they are based on reasonable assessments and are practical with respect to applications in EM field data analysis. A quantitative measure of the goodness of the least-squares solutions for a high-resolution (quasi-continuous) Z can thus be afforded. A detailed description of the rationale and implementation of the above procedure to AMT data will be reported elsewhere.

Results

The particular solution we use here for demonstration assumes that the noise exists only in the electric field components and is referred to as a downward-biased estimate.



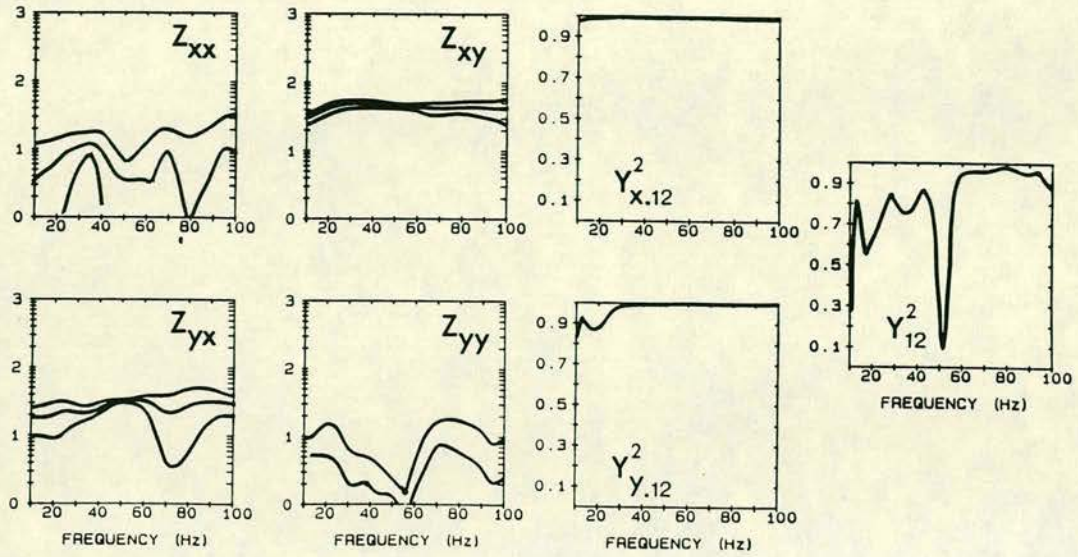
a



b

Fig. 5a–c. Three examples of the downward-biased impedance tensor elements, calculated from high-quality ELF *transient events* with the multivariate ME method. Analysis parameters are $N=50$ (0.25 s), $M=6$. 58 frequency estimates, 1.56 Hz apart. Results are the modulus of the impedance tensor, the predicted coherence functions $Y^2_{i.12}$ and the input field ordinary coherence Y^2_{12}. 1 and 2 refer to the input field components H_x and H_y, respectively. *Upper and lower curves* represent 95% confidence limits; *middle curve* is the estimated impedance

Figure 5 displays the results from three data windows containing large-amplitude ELF transient events. Figure 6 shows the equivalent results obtained from background waveforms. We present the modulus of the impedance tensor elements (logarithmic scale) as a function of frequency (linear scale), as well as the predicted coherence function ($Y^2_{i.12}$, $i=x, y$) and the input field coherence (Y^2_{12}). The subscripts x and y will henceforth refer to the predicted coherences associated with the solutions of Eqs. (1) and (2) respectively. The error bounds displayed represent the 95% confidence limits. The larger error bounds associated with the ELF transient event impedances are explained as follows. The capability of the ME method to handle short data lengths permits the analysis of such waveforms in isolation, without interference from other sources. Typically, data windows of 0.25-s (50 data points) duration have been analysed in Fig. 5. The degrees of freedom associated with the resulting spectral components are necessarily reduced



C

(e.g. $n=8$ for a filter length of $M=6$), with a corresponding widening of the confidence interval. Care should therefore be exercised when comparing the errors between the transient and the background waveform throughout this discussion; the difference may, at first glance, be misleading. In addition, the geoelectric structure of the measurement site has been identified as 2-dimensional and, by chance, the directions of the axes of measurement were almost coincident with the principal geoelectric structural directions. In such circumstances, the least-squares solutions are pathological for the diagonal elements and give rise to the instabilities observed. The data from this site permit a study of a naturally decoupled (rotated) data set. Consequently, we shall focus our attention on the off-diagonal elements.

A detailed comparison of the six results for the off-diagonal impedance pair (Z_{xy}, Z_{yx}) reveals that the estimates obtained from "single" events are consistent at the 95% confidence level. The properties which differ in the two types of waveforms analysed are the levels of predicted coherence (of both solutions) and the input field ordinary coherence. A comparison of the input field (i.e. magnetic field) ordinary coherence functions reveals that the transient events (Fig. 5) display much higher levels than do the background events of Fig. 6. The input field coherence remains large across the complete bandwidth. According to Pedersen and Svennekjaer (1984), very high Y^2_{12} may be associated with high degrees of polarization of the input field; this is certainly true for the data in question. The predicted coherence functions associated with the transient events are invariably >0.95, so that bias in the impedance estimates is expected to be minimal. The background waveforms display a high degree of stability throughout their impedance spectra. The predicted coherence is >0.80 and the input field coherence Y^2_{12} displays a strong frequency dependence and is generally much lower than that associated with the transient events. We interpret this as an effect of the data containing samples of several waveforms at various states of polarization, whose relative contribution produces the results shown. In summary, a comparison of the stable impedance elements from the two types of Schumann resonance activity indicates that they are consistent and that our estimation procedure is not influenced by polarization properties of the transient waveforms.

This assertion is examined in more detail by considering the results presented in Fig. 7. *Single*-frequency estimates for the four tensor impedance elements are shown. The frames indicated by "B" contain estimates from a number of background data windows, arranged as a function of a count number. The frames indicated by "I" contain estimates from ELF transient events, plotted as a function of their associated azimuthal angle. For this demonstration, azimuths are defined as 90° (magnetic east), 0° (magnetic north) and −90° (magnetic west). All the data windows that provided the results for Fig. 7 have been selected on the basis of high S/N ratio ($Y^2_{i,12}>0.85$, $i=x,y$). An additional requirement for the selection of the ELF transients analysed was waveform purity, i.e. absence of any other interfering sources. This has been achieved by analysing short data lengths. The real time interval within which these waveforms were collected is approximately 8 min. The estimates shown *between* the frames are averaged impedances. Averaging was performed by using the associated 95% confidence limits as weights. The first estimate from the left (marked with "B") is the average of the impedance elements in frame "B"; the third (marked with "I") is the equivalent for the impedances in frame "I". The second (unmarked) estimate is the combined result of a data set of 165 data windows with duration lengths of 0.75 s (150 data points), containing the whole range of possible Schumann resonance waveforms at various noise levels. Three example at frequencies 17 Hz, 25 Hz and 64 Hz are presented for demonstration. No systematic variation with azimuth, of the impedances derived from the linearly polarized waveforms, is evident. All estimates appear to be within the 95% limits of their population average.

In summary, it appears that the Schumann resonances can provide consistent and repeatable geoelectric sounding curves that are independent of the state of polarization of

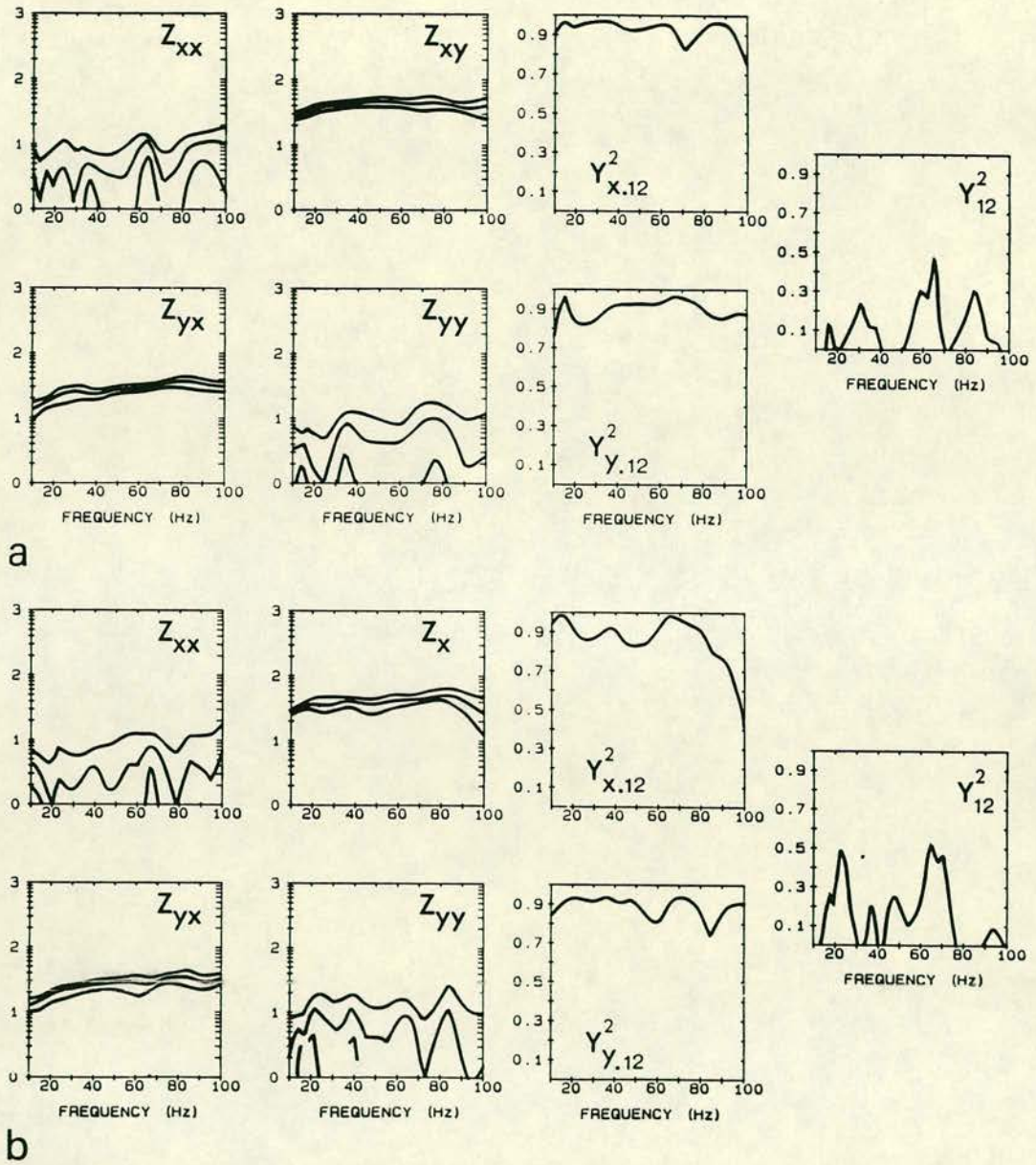


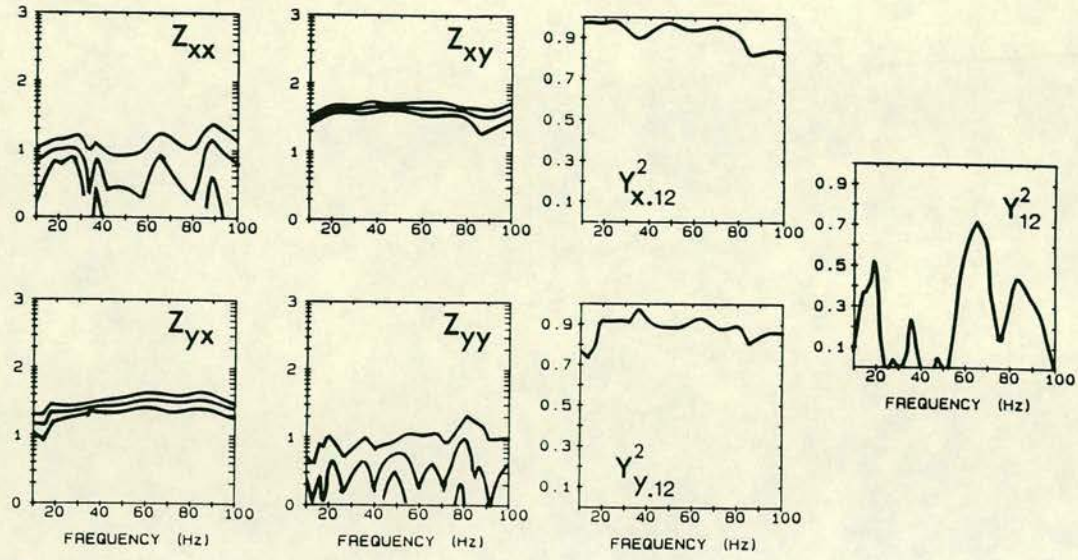
Fig. 6a–c. As for Fig. 5, but for high-quality ELF *background*. Analysis parameters are $N = 150$ (0.75 s), $M = 10$

the waveforms. Our results indicate that, ideally, geoelectric sounding can be performed with one or two high-quality data windows (e.g. the impulsive events of Fig. 4). The same conclusions can be drawn by considering the impedance phase results which have been omitted for the sake of brevity.

Summary

This study has considered Schumann resonance waveforms and their applied use for sounding crustal geoelectric structure. It has been noted that such waveforms provide the main energy component as we employ higher-frequency fields and encounter the transition from a spectrum derived from geomagnetic disturbances to a spectrum ultimately dominated by ELF sferics.

The spectral characteristics noted from previous work and the waveform polarizations investigated in this study appear consistent with theoretical models of thunderstorm energy release and propagation in the Earth-ionosphere cavity. We have examined the two main components of the energy spectrum: namely, the background field and the larger-amplitude transient waveforms. By making use of the horizontal polarization properties of the waveforms it



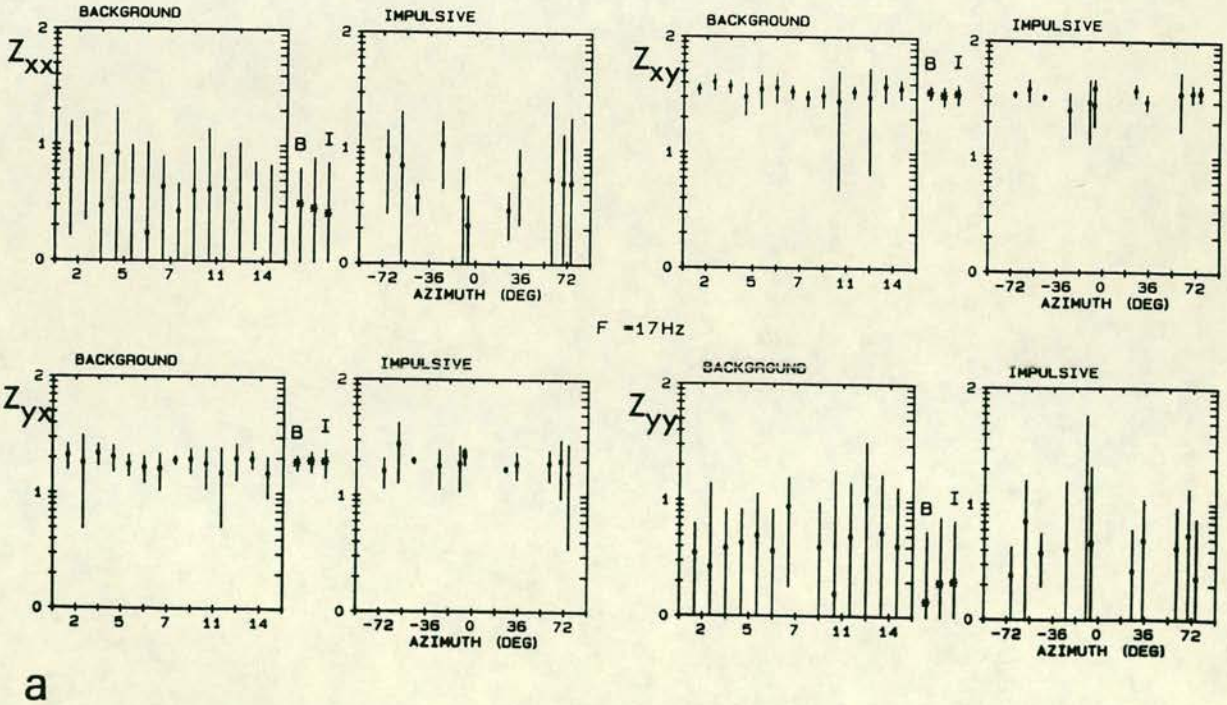


Fig. 7a–c. Comparison of impedance results from ELF background (*frame B*) and transient events (*frame I*). Three examples of *single-frequency* tensor estimates, calculated as for Fig. 6 and 5, respectively. Error bars are 95% confidence limits. **a** 17 Hz; **b** 25 Hz; **c** 64 Hz

has been possible to demonstrate the 'incoherent' nature of the background field and the arrival of larger energy components from individual thunderstorm centres distributed across the globe. By making joint use of the spectral and polarization properties of the waveforms it has been possible to obtain estimates of the range of source locations. Our data indicate that the larger energy components have propagated over large distances in the Earth-ionosphere cavity. The results, for both energy components, point toward the validity of the 'plane-wave' assumption.

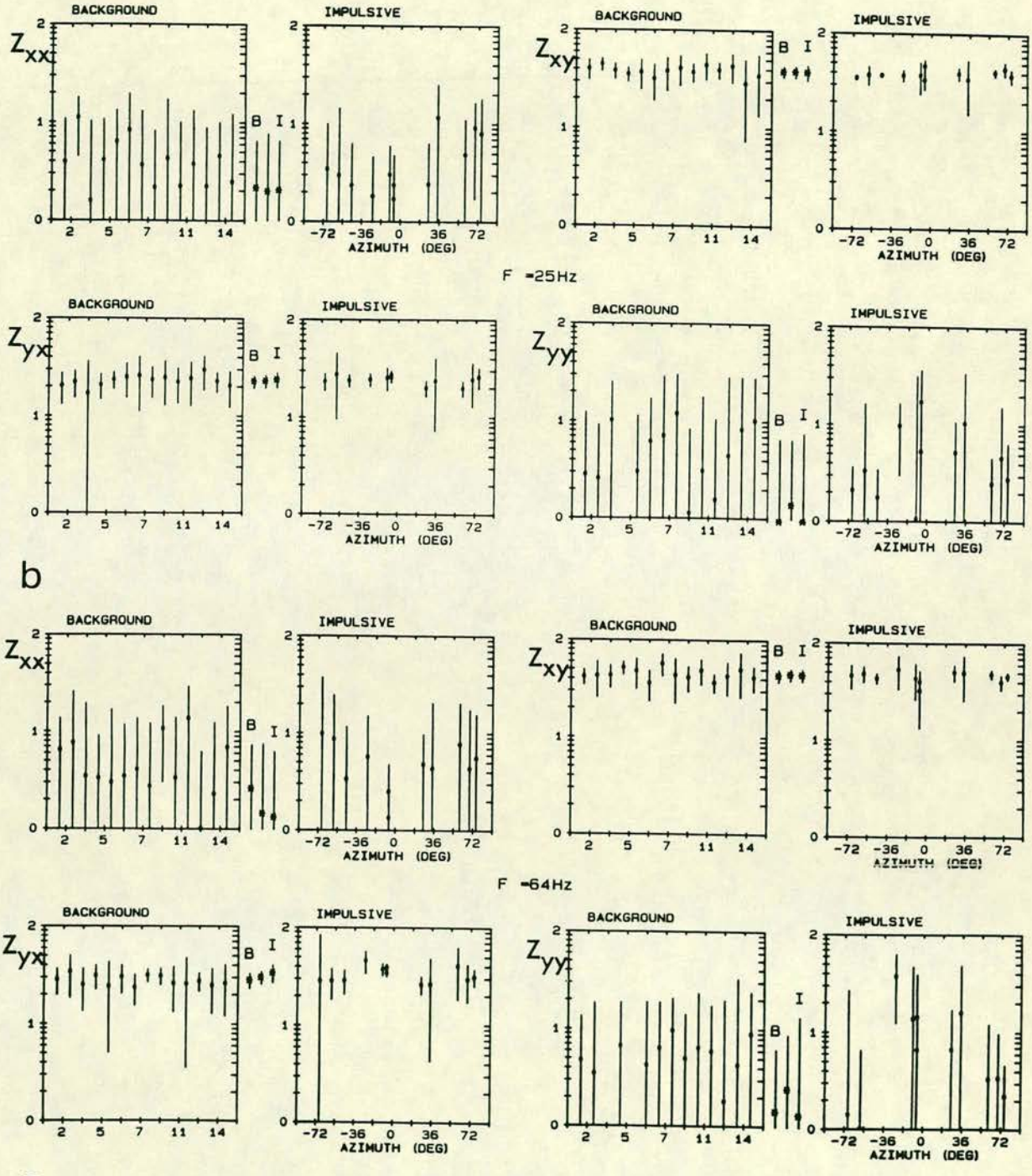


Fig. 7b, c

Our study has also considered the extent to which the Schumann resonance waveforms and their properties influence geoelectric sounding curves. In order to isolate the effects of short-duration waveforms we have found it useful to employ a maximum entropy spectral analysis scheme. Such a scheme is particularly appropriate for the transient events because it enables the geoelectric response of *individual* waveforms, directly related to a given discharge and propagation path, to be determined. Our data suggest that the two main Schumann energy components provide two

types of behaviour in the coherence functions typically determined by least-squares solutions. The background waveforms, which necessarily have the lowest S/N levels, consistently provide the lowest values in both the predicted coherence and input field ordinary coherence. When the transient events are analysed on an individual basis they are found to provide very high values in both of these coherence functions.

We have examined the behaviour of the impedance functions obtained using both types of waveform. For the background waveforms, 14 repeat determinations are shown. For the transient events 11 determinations, with the events classified according to their polarization angle, have been provided. The comparison is made for three selected frequencies. The 'individual' estimates obtained for the two types of waveform are consistent at the 95% confidence level and no systematic variation with polarization angle is evident. Using equivalent degrees of freedom, the transient events generally provide a more accurate result.

Each set of 'individual' waveform data were also combined to form 'average' results for both waveform types. The results obtained were then compared with the 'composite' results using a much larger data set comprising all waveform types. The 'individual', 'average' and 'composite' results are all entirely consistent. Our results indicate that an adequate AMT sounding for the Schumann bandwidth can be achieved using only a small number of high-quality data windows provided by the impulsive events.

Acknowledgements. The data collection was supported by the Natural Environment Research Council, the Overseas Development Administration and Kandilli Observatory, Bogazici University. This paper is published with the permission of the Director, British Geological Survey (NERC).

References

Abbas, M.: Hydromagnetic wave propagation and excitation of Schumann resonances. Planet. Space Sci. **16**, 831–844, 1968

Beamish, D., Tzanis, A.: High resolution spectral characteristics of the Earth-ionosphere cavity resonances. J. Atmos. Terr. Phys. **48**, 187–203, 1986

Bliokh, H., Nikolaenko, A.P., Filippov, Yu.F.: Schumann resonances in the Earth-ionosphere cavity (Llanwyn-Jones, D., ed.). IEE Electromagnetic Wave Series 9, P. Peregrinus Ltd. 1980

Brillinger, D.R.: Time series. Data analysis and theory. Holden-Day 1981

Burg, J.P.: A new analysis technique for time series data. Paper presented at the NATO Advanced Study Institute on Signal Processing with Emphasis on Underwater Acoustics, Enschede, The Netherlands, August 1968

Gabor, D.: Theory of communication, J. Inst. Elec. Eng. **93**, 429–457, 1946

Hermance, J.F.: Processing of magnetotelluric data, Phys. Earth Planet. Inter. **7**, 349–364, 1973

Hoover, D.B., Long, C.L., Senterfit, R.M.: Some results from audiomagnetotelluric investigations in geothermal areas. Geophysics **43**, 1501–1514, 1978

Ingmann, P., Schaefer, J., Volland, H., Schmolder, M., Manes, A.: Remote sensing of thunderstorm activity by means of VLF sferics, PAGEOPH **123**, 155–170, 1985

Jaynes, E.T.: On the rationale of maximum entropy methods. Proc. IEEE. **70**, 939, 1982

Jones, D.L., Kemp, D.T.: Experimental and theoretical observations of transient excitation of Schumann resonances. J. Atmos. Terr. Phys. **32**, 1095–1108, 1970

Keller, G.V.: Natural-field and controlled-source methods in electromagnetic exploration. Geoexploration **9**, 99–147, 1971

Keller, G.V., Frischknecht, F.C.: Electrical methods in geophysical prospecting. Pergamon Press 1966

Kemp, D.T.: The global location of large lightning discharges from single station observations of ELF disturbances in the Earth-ionosphere cavity. J. Atmos. Terr. Phys. **33**, 919–928, 1971

Kemp, D.T., Jones, D.L.: A new technique for the analysis of transient ELF electromagnetic disturbances within the Earth-ionosphere cavity. J. Atmos. Terr. Phys. **33**, 567–572, 1971

Kodera, K., Gendrin, R., De Villedary, C.: Complex representation of a polarized signal and its applications to the analysis of ULF waves. J. Geophys. Res. **82**, 1245–1255, 1977

Kromer, R.: Asymptotic properties of the autoregressive spectral estimator. Ph.D. thesis, Dept. of Statistics, Stanford Univ., Stanford, California, 1970

Labson, V.F., Becker, A., Morrison, H.F., Conti, U.: Geophysical exploration with audiofrequency natural magnetic fields. Geophysics **50**, 656–664, 1985

Lienert, B.R.: The effect of source field polarization on estimates of the magnetotelluric impedance tensor. Geophysics **45**, 1803–1812, 1980

Madden, T., Thomson, W.: Low-frequency electromagnetic oscillations of the Earth-ionosphere cavity. Rev. Geophys. **3**, 211–254, 1965

Morf, M., Vieira, A., Lee, D.T.L., Kailath, T.: Recursive multichannel maximum entropy spectral estimation. IEEE Trans. on Geoscience Electronics **GE-16**, 85–94, 1978

Pedersen, L.B.: The magnetotelluric impedance tensor – its random and bias errors. Geophys. Prospect. **30**, 188–210, 1982

Pedersen, L.B., Svennekjaer, M.: Extremal bias coupling in magnetotellurics. Geophysics **49**, 1968–1978, 1984

Priestley, M.B.: Spectral analysis and time series – Volume 2: multivariate series, prediction and control. Academic Press 1981

Russell, C.T., Holzer, R.E., Smith, E.J.: OGO 3 observation of ELF noise in magnetosphere, II. The nature of equatorial noise. J. Geophys. Res **75**, 755, 1970

Sims, W.S., Bostick Jr, F.X., Smith, H.W.: The estimation of magnetotelluric impedance tensor elements from measured data. Geophysics **36**, 938–942, 1971

Slankis, J.A., Telford, W.M., Becker, A.: 8-Hz telluric and magnetotelluric prospecting. Geophysics **37**, 862–878, 1972

Strand, O.N.: Multichannel complex maximum entropy (autoregressive) spectral analysis, IEEE Trans. on Autom. Control **AC-22**, 634–640, 1977

Strangway, D.W., Swift Jr, C.M., Holmer, R.C.: An application of audio frequency magnetotellurics (AMT) to mineral exploration. Geophysics **38**, 1159–1175, 1973

Telford, W.M.: Characteristics of audio and sub-audio telluric signals. Geophys. Prospect. **25**, 321–333, 1977

Tzanis, A., Beamish, D.: Time domain polarization analysis of Schumann resonance waveforms. J. Atmos. Terr. Phys. 1986a (in press)

Tzanis, A., Beamish, D.: Observations of the polarization properties of sferic waveforms. GRG Internal Report 86/13, British Geological Survey, 1986b

Ulrych, T.J., Bishop, T.N.: Maximum entropy spectral analysis and autoregressive decomposition. Rev. Geophys. Space Phys. **13**, 183–200, 1975

Uman, M.A.: Lightning, McGraw-Hill 1969

Ville, J.: Theorie et applications de la notion de signal analytique. Cables Transm. **1**, 61–74, 1946

Vozoff, K.: The magnetotelluric method in the exploration of sedimentary basins. Geophysics **37**, 98–114, 1972

Watt, A.D.: VLF radio engineering. Pergamon Press 1967

Received November 11, 1986; revised version February 5, 1987
Accepted February 13, 1987